Essays in Empirical Political Economics

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
der Universität Mannheim

Andreas Bernecker

vorgelegt im Frühjahr 2014

Abteilungssprecher: Professor Dr. Eckhard Janeba

Referent: Professor Dr. Eckhard Janeba

Korreferentin: Professor Christina Gathmann, Ph.D.

Tag der mündlichen Prüfung: 8. Juli 2014

Acknowledgments

First and foremost, I thank Eckhard Janeba and Christina Gathmann for their invaluable support and constant advice. I felt truly privileged to benefit from such great advisors. I also thank Christina Gathmann for being such a pleasant and inspiring co-author to work with on Chapter 3 of this dissertation. For additional guidance and help, I want to thank Hans Peter Grüner, Pierre Boyer, and Panu Poutvaara.

I further thank my colleagues at the University of Mannheim, in particular my fellow students at the Center for Doctoral Studies in Economics and at the Chair of Public Finance and Economic Policy. Special thanks go to my office mates Alexander Paul and Hannes Kammerer and to Christian Koch for numerous coffee sessions.

For jointly making it past the first year of graduate studies, I thank my fellow students at the University of California, Berkeley, in particular Santiago Pereda Fernandez, Tadeja Gracner, Tarso Mori Madeira, Sinaia Urrusti Frenk, and Paolo Zacchia.

I also thank the Deutsche Forschungsgemeinschaft and the German Academic Exchange Service for financial support and colleagues at several other institutions for their hospitality during my visits there. In particular, I want to name the Ifo Institute Munich, Stockholm University, the Stockholm School of Economics, the Uppsala Center for Fiscal Studies, the University of Konstanz, and the European Central Bank.

Finally, I want to thank my parents Joachim and Gabi, my siblings Kerstin and Nicolai, and – last but definitely not least – Vincent. Without their backing and unconditional love this project would have been condemned to failure.

Mannheim, Spring 2014

Andreas Bernecker

Contents

List of Figures

List of Tables

Chapter 1

General Introduction

Economists tend to have many good ideas on how to improve people's lives. However, they are frequently disappointed that these ideas are often not implemented by politicians. The aim of political economics as a field therefore is to employ methods from economics to better understand political decisions. This dissertation consists of four self-contained chapters focusing on different topics within empirical political economics. While the political economics of reforms can be considered the connecting theme for the most part of this dissertation, the different chapters analyze governments, executive political actors, legislative actors, and finally voters themselves one after another.

In Chapter 2, I investigate the relationship between divided government and the adoption of economic reforms at the state level during the US Welfare Reform. According to long-standing theories, governments where the executive is dominated by another party than the legislative block themselves and pass fewer economic reforms. I show in this chapter that, at least for welfare reforms in US states between 1978 and 2010, the contrary is the case. An earlier version of this chapter has been circulated in the CESifo Working Paper Series (Bernecker (2014)).

In Chapter 3, which is joint work with Christina Gathmann, we analyze the role of reelection concerns of state governors during the US Welfare Reform in shaping state-level policy reform decisions. Our empirical analysis is guided by theoretical considerations and shows that reelection concerns of the governor are indeed a key driver of the decision whether to politically experiment with a new welfare policy rule or to revert a previous experiment. An earlier version of this chapter has been circulated in the University of Mannheim Economics Department Working Paper Series (Bernecker and Gathmann (2013)).

In Chapter 4, I test whether German members of parliament adjust their behavior in parliament depending on the level of political competition they have to face in their districts.

For members of opposition parties, the results show that politicians who expect close races for themselves in upcoming elections have lower absence rates in parliament. An earlier version of this chapter has been circulated in the University of Mannheim Economics Department Working Paper Series (Bernecker (2013)).

In Chapter 5, I analyze why people tend to overly stick to the status quo. Is it due to psychological processes or economic reasoning? Using experimental surveys, I show that anchoring as a cognitive misperception can solely cause one half of the whole phenomenon called status quo bias. Economic reasoning thus presumably is the force of minor power here.

Additional materials for each chapter such as data descriptions and sources, experimental surveys, and additional results are contained in a joint appendix of the dissertation. The bibliography at the end jointly lists all chapters' references.

Before introducing Chapters 2 to 5 in more detail, let me clarify the relationship of Chapter 4 to my previous work. Chapter 4 of this dissertation builds on my master thesis (which served as a dissertation proposal at the same time) written at the University of Mannheim in 2011 (Bernecker (2011)). Several sections of Chapter 4 such as the introduction or the institutional background section largely include similar or identical parts compared to my previous work. There are, however, relevant differences compared to Bernecker (2011). Let me shortly mention the four most important ones here. First, several sections such as the literature and the results section have been completely rewritten and improved. Second, the empirical approach and the regression analyses have been revised. In particular, the instrumental variable strategy is now much more reliable. It now includes instruments from further back in the past and takes into account redistricting of constituencies. Third, the analysis of politicians' outside incomes included in Bernecker (2011) has been dropped completely since from today's perspective these data seem too unreliable. Fourth, several new analyses are now included: Additional data collected from a transparency website now allow constructing a measure of politicians' motivation. New data on politicians elected to parliament via party lists is employed. This yields new results extending the analysis beyond district politicians for the first time. Chapter 4 also additionally covers politicians from the Christian Social Democrats (also not included in Bernecker (2011)).

1.1 Divided We Reform? Evidence from US Welfare Policies

More than one half of all US state governments are divided, i.e. the executive power is dominated by another party than the legislative power. Standard theory suggests that political reforms are less likely to be passed under divided as opposed to unified government. For example, the veto player theory states that the adoption of a reform is the more unlikely the more parties have a say in policy-making since the parties may block each others' reform initiatives (Tsebelis (1995), Tsebelis (2002)). In contrast to that theory, I suggest that divided government may also result in policy competition between the relevant players and thus in more reforms being passed. Whether divided governments reform more or less than unified governments therefore is an empirical question. In Chapter 2, I try to answer this question for the US states.

For that purpose, I use a rich novel data set on welfare reforms implemented at the US state level between 1978 and 2010 during the US Welfare Reform. The US Welfare Reform marks the largest shift in welfare policy in the US since the New Deal. Welfare reforms are particularly interesting to look at since welfare politics is an area at the center of public economics and a large share of public funds is typically spent on it. Besides, the welfare system in the US has undergone dramatic changes during the time under consideration. This makes welfare a relevant topic to be empirically analyzed also in its own right. The data set is complemented by ample information on US states regarding politics, ideology, and socioeconomic development.

The analysis shows that divided governments are in fact between 20% and 50% more likely to reform the welfare system than unified governments. This result is robust with respect to numerous checks such as the inclusion of year and state fixed effects, adding various control variables, employing different estimation techniques, controlling for spillovers across states, or changing the definition of what is considered to be a welfare reform. The effect also stays sizable and significant when focusing the analysis on close elections providing quasi-random variation in the style of a regression discontinuity design. Contrary to conventional wisdom, divided US state governments reform more than unified ones.

The empirical evidence is consistent with an explaining theory that is based on policy competition between the state governor, the state senate, and the state house. According to the theory, policy competition which is fruitful for passing economic reforms is more intense under divided than under unified government. This theory is also supported by a range of welfare reform case studies. Other potential explanations on the other hand are not consistent with the data.

Thus, it seems that voters in US states should not be too afraid of potentially getting into reform deadlock when electing a divided government. And although Chapter 2 only analyzes reforms at the US state level, the results at least call into question typical reform deadlock arguments frequently made for other countries as well. In fact, governments that are not completely dominated by one party may even be beneficial for passing economic reforms.

1.2 Trial and Error? Reelection Concerns and Policy Experimentation during the US Welfare Reform

As outlined before, not enough is known about how the political process shapes political reforms. What are the political incentives driving policy experiments and learning from past experiments? While Chapter 2 looks at the effect of different forms of government on reform adoption, Chapter 3 instead focuses on the head of the executive as the key policy maker in the spot light. Based on a political agency model of reputational concerns, we derive empirical predictions that are tested using the same welfare reform data set as before. The specific question to be answered is: How do reelection concerns of the governors affect experimenting with new welfare policies and reverting of previous experiments in US states during the US Welfare Reform?

The model of reputational concerns we line out builds on the assumption that experimenting with new policies reveals information to the governor about the appropriateness of a new policy. However, it also reveals information to the voters about the quality of the governor as a lawmaker. In that case, a governor may have incentives to experiment strategically to signal quality to voters in order to increase her or his reelection chances. The model consists of two periods. In the first period, the governor either implements a new policy or sticks to the status quo policy. Then, pay-offs from the policy are realized. In the second period, in case the governor decided to reform in the first period, she or he chooses whether to stick to the new policy or to revert to the status quo policy. Again, pay-offs are realized. Voters observe all the governor's decisions and its pay-offs. They use this information to update their beliefs regarding the governor being of high or low quality. Based on these beliefs, they finally decide whether to reelect the governor or not.

We derive three empirical predictions from that model of reputational concerns: First, low-quality governors experiment less when the expected welfare pay-offs from a reform experiment for society are high. The intuition is the following: Only being successful with an experiment with ex ante *low* expected pay-offs can boost a low-quality governor's reputation enough to be mistakenly taken to be a high-quality governor by voters. Second, a low-quality governor is less likely to reverse a reform experiment that failed in the first period if ex ante the expected pay-offs from the experiment were high. The reason is the large reputational boost for the governor in the unlikely case the experiment turns out to be successful in the second period. Third, the higher the initial reputation of a governor the less likely she is to experiment. The intuition is that such a governor has to lose much in terms of reputation in case an experiment fails.

We find relevant and significant evidence for all predictions in the welfare reform data: First, a one standard deviation increase in governor quality and the expected pay-off from experimentation raises the likelihood of an experiment by between 37 and 49%. Second, among low-quality governors, a one standard deviation increase in the expected pay-off reduces the likelihood of a reversal by 36%. Third, a one standard deviation increase in initial reputation reduces the likelihood of an experiment by 27%. A large set of different robustness checks confirms these findings. We also find strong evidence that governors who cannot be reelected, and should therefore be less concerned about their reputation among voters, indeed behave systematically differently with respect to experimenting with and reverting reforms. Consistent with our theory, not only the form of government (Chapter 2), but also electoral incentives of key policy makers are thus highly relevant for understanding the adoption of economic reforms.

1.3 Do Politicians Shirk when Reelection Is Certain? Evidence from the German Parliament

Chapter 4 switches the focus to legislative actors. Instead of working in parliament, legislators may sometimes shirk and invest their time in activities outside parliament. As in Chapter 3, elections are considered the main device for voters to keep politicians accountable. The specific question asked in Chapter 4 is: Does stiffer political competition via elections reduce shirking by members of parliament? For a micro-analysis of this issue, I construct and analyze a new data set covering German members of parliament. The hypothesis is that a legislator elected in her district with a very large vote margin ahead of the opponent can almost surely expect to be reelected in the next election even if her working attitude in parliament is not exemplary. On the other hand, a legislator who was barely elected in her district faces a considerably higher risk of losing reelection if she does not adjust her behavior in parliament accordingly. Thus, I expect small vote margins to lead to lower shirking at the legislator level.

The data set I construct to test this hypothesis spans the years 2005 to 2012 and covers a range of biographical and political information on the individual legislator level. Shirking is operationalized by absence rates at mandatory parliamentary sessions. To be able to control for special circumstances, the data set also comprises information on excuses (e.g. due to health issues), on special functions of the legislators (such as being faction leader or government minister), and further data. The baseline analysis is done by ordinary least squares with the absence rate as the dependent and the vote margin as the main independent variable. Since ordinary least squares estimates could however be biased, e.g. due to highly motivated legislators having large vote margins and low absence rates at the same time, they are complemented by two-stage least squares estimates. The instrument used for the candidate's individual vote margin relies on convenient peculiarities of the German electoral system and is based on the votes a candidate's party gets.

The results show that indeed legislators with small vote margins have significantly and relevantly lower absence rates in parliament. A more detailed analysis reveals that this effect is particularly strong for legislators from opposition parties while it is insignificant for legislators of governing parties. The explanation for this finding could be that in governing parties the faction leaders enforce attendance in parliament such that the government can get its bills passed. For legislators of opposition parties, an increase in the vote margin by 10 percentage points increases the absence rate by about 7 percentage points. The results are robust across different specifications. One can also show that legislators with high absence rates are indeed punished by voters in subsequent elections. Furthermore, the chapter extends the analysis to legislators elected via party lists showing that a safer list position results in a significantly higher absence rate. This is consistent with the finding for district legislators: Political competition has a disciplining effect on legislators.

Thus, political competition does not only affect politics via the form of government (Chapter 2), but also has direct effects on individual executive (Chapter 3) and legislative actors (Chapter 4). Chapter 5, finally, analyzes voters themselves.

1.4 Is Status Quo Bias Explained by Anchoring? Evidence from Survey Experiments

In a seminal paper, Samuelson and Zeckhauser (1988) have shown that people stick to the status quo more frequently than predicted by standard economic theory and labeled this phenomenon "status quo bias". Ample examples of it can be found in everyday situations such as returning to the same vacation spot each year. But also in politics status quo bias is ubiquitous. This often results in complaints that no economic reforms can be passed since voters do not like change. Samuelson and Zeckhauser (1988) offer different possible explanations for status quo bias in people's preferences. They categorize them into rational ones based on economic reasoning such as transition cost on the one hand and cognitive misperceptions such as anchoring on the other hand. Chapter 5 tries to make a first step towards empirically disentangling these different explanations. In particular, I aim to show how quantitatively relevant anchoring is relative to other explanations of status quo bias.

Anchoring is a cognitive misperception that leads people to make judgments that are biased towards initially presented values. One of the best illustrations of this effect is given by Ariely et al. (2003). They ask participants in an experiment to state the last two digits of their social security number. Then they offer participants a bottle of wine and ask them if they would be willing to pay for the wine more or less than the number they just stated in dollars. In a follow-up question, they also ask for the exact willingness to pay for the wine. Ariely et al. (2003) find that people's social security number statement and their stated willingness to pay for the wine are in fact highly correlated: People with a high number are willing to pay more for the wine (and vice versa). This example illustrates how powerful anchoring to random numbers can be. Due to its strong effects, anchoring is one of the preferred explanations for status quo bias in the eyes of Samuelson and Zeckhauser (1988).

For disentangling anchoring from other potential causes of status quo bias, I run large survey experiments. The design is split sample, i.e. different treatment groups of respondents of the survey get different versions of the survey. Using different treatments such as a status quo effect treatment following Samuelson and Zeckhauser (1988), an anchoring treatment following Ariely et al. (2003), and other control treatments allows me to estimate how much of status quo bias is due to anchoring. The surveys have been implemented in the representative German Internet Panel (GIP) and in a classroom experiment at the University of Mannheim. In total, more than 1,500 respondents participated.

The results suggest that anchoring as a cognitive misperception (opposed to explanations based on economic reasoning) can alone cause one half of the status quo bias phenomenon. It is difficult to draw definitive policy conclusions from this. However, some potential take-aways come to mind. First, it seems a good idea not to over-interpret status quo bias in voters' stated policy preferences. In fact, it could be a worthwhile endeavor in some cases to convince voters that more economic reforms may be necessary. Second, surveys trying to reveal policy preferences should in general be interpreted carefully. When designing future surveys, anchoring as a strong empirical phenomenon potentially falsifying results should definitely be taken into account. Third, researchers in particular should be more careful when presenting numbers to people somewhere as part of their research design. Learning more about the anchoring effect's interactions with other phenomena can inform us about the validity of parts of the research in experimental economics and about how future research designs should be optimized.

Chapter 2

Divided We Reform? Evidence from US Welfare Policies[1]

2.1 Introduction

"Now, hug a Republican", the Economist told President Obama via the title of its November 10th issue after he had won reelection in 2012 (The Economist (2012)). The newspaper referred to the fact that Democrat Obama would again have to deal with a Republican majority in the House of Representatives. As before the election, government would be divided. Divided government means that the President is faced with a majority of another party in at least one of the two chambers of Congress. Usually, it is argued that this hinders legislative productivity since the government cannot get its bill proposals through Congress without getting the consent of the opposition party. The legislative majority may even decide to block any relevant initiatives

[1]I thank the International Institute of Public Finance for awarding an earlier version of this chapter the "Peggy and Richard Musgrave Prize" for the best paper presented by an economist under the age of 40. For helpful comments and discussions, I thank Thushy Baskaran, Johannes Becker, Serra Boranbay, Pierre Boyer, Thomas Bräuninger, Ralf Brüggemann, Micael Castanheira, Antonio Ciccone, Valentina Corradi, Marc Debus, Georgy Egorov, Tore Ellingsen, Olle Folke, Ronny Freier, Clemens Fuest, Vincenzo Galasso, Christina Gathmann, Hans Peter Grüner, Mark Hallerberg, Zohal Hessami, Björn Tyrefors Hinnerich, Eckhard Janeba, Magnus Johannesson, Leo Kaas, Georg Kirchsteiger, Thomas König, Miklos Koren, Christian Lessmann, Erik Lindqvist, Johannes Lindvall, Alessandro Lizzeri, James Lo, Andreas Madestam, Francois Maniquet, David Mayhew, Claudio Michelacci, Sten Nyberg, John Nye, Torsten Persson, Per Pettersson-Lidbom, Justin Phillips, James Poterba, Niklas Potrafke, Panu Poutvaara, Sven-Oliver Proksch, Steffen Reinhold, Johanna Rickne, Mark Schelker, Mary Shirley, Albert Solé-Ollé, Thomas Stratmann, David Strömberg, Carsten Trenkler, George Tsebelis, Heinrich Ursprung, Daniel Waldenström, Andrea Weber, Richard van Weelden, Galina Zudenkova, and many others. I also thank seminar participants at University of Mannheim, Stockholm University, Stockholm School of Economics, European Central Bank, Uppsala Center for Fiscal Studies, Lund University, University of Konstanz, Ifo Institute Munich, Centre for European Economic Research, and at the following conferences and workshops: European Public Choice Society Meeting, Institutions and Politicians Conference at Centre for European Economic Research, Spring Meeting of Young Economists, International Society of New Institutional Economics Conference, International Institute of Public Finance Annual Meeting, European Economic Association Meeting, European Winter Meeting of the Econometric Society, CESifo Workshop on Political Economy. An earlier version of this chapter has been circulated in the CESifo Working Paper Series (Bernecker (2014)).

taken by the President resulting in complete legislative deadlock. Similar deadlock arguments are often also made with respect to comparable situations of partisan divide in other countries. This paper systematically analyzes this issue for the US states level by answering the following question: Is it really true that actual political reforms are less likely under divided as opposed to unified government? – I show that the contrary is in fact the case.

The standard deadlock argument made with respect to divided government is that differing partisan dominance of different institutions leads to a lower propensity to reform since the different parties have to agree on how to deviate from the status quo. This intuition has been theoretically formalized by George Tsebelis in his seminal work on veto players (Tsebelis (1995), Tsebelis (2002)): The more veto players have a say in policymaking, the less likely are reforms changing the status quo. Similarly, Howitt and Wintrobe (1995) show in a theoretical model how political inaction may result when both parties have power and competition is stiff. Along with conventional wisdom, theory thus clearly predicts that one should expect fewer reforms under divided government compared to unified government. But is this the whole story? – Maybe different party dominance of different governmental bodies enhances policy competition between them leading to more reforms in the end? This is why this paper sets out to empirically assess reform adoption by divided versus unified governments.

I investigate whether welfare policies are more likely to be reformed under divided or under unified government using novel data from the US states level from 1978 to 2010. During this period of time, more than one half of all US state governments were divided. Welfare politics is an interesting case to look at for at least three reasons. First, welfare is one of the policy areas most central to economics and also among the largest budget items both at the US federal level and at the state level (US Government Spending (2013)). Second, during the time span analyzed in this paper the US Welfare Reform was at the center of the public debate since it represented the largest shift in welfare politics since the New Deal in the 1930s. US states reformed important elements of the welfare system such as work requirements, sanctions, and time limits. However, despite the large public and political interest in these reforms and a large policy evaluation literature on the topic, the political economy aspect is heavily underresearched. Third, welfare politics is a perfect field for the analysis of the effects of divided government since, along with the governors, state legislatures played a key role in the process. If divided government indeed leads to political parties blocking each other, one should definitely observe this for a very partisan issue such as welfare politics. Given all this, this paper analyzes a rich data set on welfare reforms at the US state level constructed from several different sources. Welfare policy changes for all US states are coded on a yearly basis and a wide range of demographic, political, and ideology controls are included. The resulting novel data set gives a comprehensive overview of welfare reform activity in US states between 1978 and 2010.

Different measures of welfare reform are then used as dependent variables in panel data regressions. The main explanatory variable is divided government. By divided government I mean that the state governor is confronted with a majority of legislators of the other party in one or both of the chambers of the state legislature. Including fixed effects allows within-state identification, i.e. the analysis compares the reform effects of a unified government in Wisconsin to a divided government in Wisconsin (and not a unified government in Wisconsin to a divided government in New York). I show that under divided government a US state's probability to implement a welfare reform is actually between 5 and 10 percentage points *higher* than under unified government. The size of this effect amounts to between 20 and 50% of the unconditional probability of a US state to implement a welfare reform between 1978 and 2010. The effect is highly significant and stable across specifications. This result is in stark contrast to conventional wisdom and standard theory.

To check the robustness of my finding, I explore potential issues of omitted variable bias, measurement of the dependent variable, reverse causality, differential treatment effects, and estimation technique. To avoid omitted variable bias, I test the inclusion of standard demographic controls, variables related to welfare reform and welfare state crisis, political and ideological controls. I also include year fixed effects and control for a state's reform history and reform spillovers across states. None of these controls affects the result. I also show that the effect does not depend on the way welfare reform is measured in the data and does also hold when considering only large reforms, for example. To take care of reverse causality and further potential endogeneity concerns, I focus on close elections providing quasi-random variation in the type of government. Still, I find the positive and significant effect of divided government on reform adoption. An analysis of different subsamples allows investigating differences in the effect, e.g. for Southern versus non-Southern states. Finally, also with respect to employing different estimation techniques the result is very robust. Divided governments are more likely to reform than unified governments.

Why do divided governments reform more? – I suggest policy competition between governor, senate, and house as potential explanation. It is well known that these three institutions typically engage in stiff competition with each other (e.g. Rosenthal (2009), p. 197). Under different partisan dominance this competition may be even more intense. The relevant difference between unified and divided government is that under the latter also the opposition party has agenda setting power: By passing bills the opposition party dominating a legislative chamber can confront the governor with policy issues. Policy competition between the actors may arise and more innovative policies may be implemented in the end. In some cases, the opposition leader may even want to qualify as able future governor in the eyes of the voters by passing innovative reform bills.

An empirical analysis based on different types of divided government provides evidence that is consistent with the policy competition theory. It is also supported by welfare reform case studies. For example, in Wisconsin Democrats having the majority in the state legislature suggested even more drastic welfare reforms than Republican governor Tommy Thompson who is known as a very ambitious reformer of welfare (e.g. Wiseman (1996), p. 532). And in New Jersey, Democratic assembly majority leader Wayne R. Bryant – and not the governor – was the main mover of welfare reform (e.g. Haskins (2006), p. 34). Both in Wisconsin and in New Jersey government was divided at that time. I also empirically explore several other potential explanations of my finding, but none of them is consistent with the data. It indeed seems to be competition between governor, senate, and house that makes divided government reform more. Finally, the passage of the US Welfare Reform at the federal level in 1996 is itself an illustrative example: The Republicans used their majority in both chambers of Congress to challenge Democratic President Clinton in the realms of welfare reform by passing significant reform bills. The Clinton administration reacted using vetoes and counter-proposals. In the end, the largest welfare reform since the New Deal was adopted under divided government.

The following section presents the related literature. Section 2.3 gives some background on US welfare politics and presents the data. Section 2.4 covers the estimation strategy and results. Section 2.5 explores potential explanations of the counter-intuitive finding. Section 2.6 concludes.

2.2 Related Literature

My work relates to the growing strand of literature on causes and consequences of divided government. Classics on the causes include, for example, Alesina and Rosenthal (1995), Alesina and Rosenthal (1996), and Alesina and Rosenthal (2000) who put forward a balancing theory of divided government, i.e. voters split political power between political actors of different partisanship to get an ideologically intermediate policy in the end.[2] Another classic is Chari

[2]See Fiorina (1996) for a similar argument and an overview of more classical arguments.

et al. (1997) arguing that divided government may be a result of voters wanting a Republican who is good at keeping overall taxes down as president, but a Democrat who is good at bringing pork home as constituency representative in Congress.[3] More recent work stresses the control element of divided government: While Fox and Weelden (2010) present more effective overseight as a theoretical argument in favor of having a divided government, Schelker (2012) shows that voters – to restrict power of the unaccountable – are indeed 10% more likely to elect a divided government into office when the incumbent governor cannot be reelected.

The literature on consequences of divided government has so far mainly focused on fiscal policy and budgets. Roubini and Sachs (1989) is an early paper showing for OECD countries that during crises coalition and minority governments in general are bad at managing the budget. For US states in particular, Poterba (1994) and Alt and Lowry (1994) show that unified governments are better able to respond to fiscal crises. More recent work stresses how budgets tend to be late under divided governments, see Klarner et al. (2012) and Andersen et al. (2012). The latter paper, for example, finds that the budget is 10 to 20% more likely to be late under divided government and offers a theoretical explanation based on a war of attrition between the parties (Alesina and Drazen (1991)).

In contrast, the present paper is concerned with the effect of divided government on the adoption of economic reforms. Most theoretical work has focused on reform deadlock as a natural consequence of divided government. Classics on this are Sundquist (1988) and Cutler (1988). More recent examples include Howitt and Wintrobe (1995), Tsebelis (1995), and Tsebelis (2002). The former make the theoretical argument that under stiff political competition such as under divided government no party may dare to bring up a political issue since it is afraid that the opposing second party may be strong enough to implement a policy that is even worse than the status quo from the perspective of the first party. The potential result is reform inaction. Similarly, the work by Tsebelis shows that the likelihood of reforms drops in the number of relevant political veto players and the partisan diversity among them. Again, reform deadlock is the predicted consequence of divided government.

On the empirical side, however, evidence is scarce. There is a literature on policy innovation in political science started by Walker (1969) and reviewed in Berry and Berry (2007).[4] Important examples analyzing the effects of divided government on legislative accomplishment at the US federal level are Mayhew (2005) and Binder (1999). The latter is extended in Binder (2003).

[3] See Jacobson (1990) for a related argument.

[4] Berry and Berry (1990) and Berry and Berry (1992) are important examples using event history analysis that both touch the topic of divided government.

While Mayhew shows that the number of passed laws that are relevant according to expert judgment does not differ between divided and unified governments, Binder argues that one has to take the overall political agenda into account. She measures legislative accomplishment as the share of bills passed out of all bills discussed in newspaper editorials and finds that divided governments gridlock on a larger share of the agenda. Shipan (2006) has however shown that the latter approach has the problem that the political agenda tends to be larger under divided governments and is thus endogenous. A recent overview of this inconclusive literature is provided in Binder (2011). Results depend to a large degree on the definition of legislative productivity and the focus is almost exclusively on the US federal level giving not more than 30 observations for the analysis.[5] In contrast, this paper investigates the US state level allowing for considerably deeper econometric analysis and looks at de facto implemented economic reforms in the well-defined policy area of welfare reform.

So far, there is no literature in economics analyzing the effect of divided government on the adoption of particular economic reforms. Bjørnskov and Potrafke (2013) analyze how party ideology in US states affects an economic freedom index containing, for example, tax revenue as share of GDP and union density. Although the authors interact their party ideology measures with different forms of government, the main interest lies in the effect of ideology in this paper. Castanheira et al. (2012) look at tax reforms in particular, but investigate a cross-section of European countries with different political systems. They find that countries with more parties in government are more likely to reform their tax system.

The paper also contributes to the literature on the US Welfare Reform. There is a large policy evaluation literature on this issue. For reviews, see Harvey et al. (2000) or Blank (2002). Most of this literature does, however, not take into account the political economy aspect of the reform: Since it is not random which states reform their welfare system, the evaluation literature may be faced with problems of endogeneity. To better understand which states reform and why can therefore also help to better understand the effect of different welfare policies. There is almost no work on the political economy of the US Welfare Reform. Some examples focusing mostly on welfare caseload, racial issues, and interactions between federal and state level as explanatory factors are Lieberman and Shaw (2000), Soss et al. (2001), Fellowes and Rowe (2004), and Soss et al. (2008). Bernecker and Gathmann (2013) look at the relevance of reputational concerns of US governors in shaping the US Welfare Reform. None of these papers investigates the effect of

[5] Bowling and Ferguson (2001) and Rogers (2005) are exceptions looking at the state level. But the former conducts a cross-sectional analysis of the 1994 legislative sessions only and the latter a cross-sectional analysis of 23 states only. Besides being unable to run within-state analyses, both papers do not take potential endogeneity of divided government into account.

divided government.

In broader terms, this paper is part of the literature in economics analyzing the political economy of reforms. Most of this literature is theoretical. Famous examples are Fernandez and Rodrik (1991), Dewatripont and Roland (1995), or Cukierman and Tommasi (1998). In terms of methods, the paper is closest to the political economy literature analyzing policy choices in US states (often using panel data regressions). Important examples include Besley and Case (1995a), List and Sturm (2006), and Besley et al. (2010). For an overview, see Besley and Case (2003). None of these looks at divided government or welfare reforms in particular.

2.3 Background and Data

2.3.1 US Welfare Politics

Before the landmark US Welfare Reform under President Clinton in 1996, the "Aid to Families with Dependent Children (AFDC)" program had been in place for several decades. As an entitlement program, it provided financial assistance to eligible families and almost all of its rules were determined at the federal level. Since 1962, states had the possibility to apply for welfare waivers at the Department of Health and Human Services at the federal level under Section 1115 of the Social Security Act. If approved, states could deviate from the rules set at the federal level and experiment with own policy rules as suggested in the waiver application. Such waivers became common in the 1980s when welfare caseloads began to rise and many states wanted to restrict welfare (Lieberman and Shaw (2000)). The common spirit of many such waivers was to go "from welfare to workfare". Major policy changes implemented include work requirements, family caps, time limits, and sanctions. See Harvey et al. (2000) for more details on these waivers.

In 1996, President Clinton signed the "Personal Responsibility and Work Opportunity Act" which abolished the "Aid to Families with Dependent Children (AFDC)" in favor of the new "Temporary Assistance for Needy Families (TANF)" program with new federal rules. Within these federal guidlines, the reform also granted states more liberty to decide on their own welfare policy rules and in fact decentralized welfare to the state level. Now, states no longer have to apply at the federal level when they want to reform the welfare system. The 1996 federal level reform also further strengthened states' financial incentives to reform welfare by switching from matching to block grants. Policy changes in the areas of work requirements, family caps, time limits, and sanctions have remained popular at the state level until today.

The 1996 US Welfare Reform is usually considered the most important one since the New Deal. Still, the political economy of it seems heavily underresearched. We know almost nothing about which states decided to reform their welfare systems and why. Welfare reform case studies and anecdotal evidence suggest that the governors and their electoral concerns play a very important role. This is analyzed in detail in Bernecker and Gathmann (2013). But since no welfare legislation can be passed without the consent of state senate and state house, the state legislatures clearly also have their parts. Both, governors and state legislators have been identified as "key actors" in the process of welfare reform (Liebschutz (2000), p. 18).

In many states, reforming governors intensively collaborated or struggled with their state legislatures. Liebschutz (2000), for instance, gives examples from Florida, Mississippi, New York, Washington, and Wisconsin (pp. 19, 60, 109). In several of these instances government was divided. There is also evidence that in some states it was the state legislatures even taking the initiative in the welfare reform process, for example in Wisconsin in 1979 (before well-known reformer Tommy Thomspon took the gubernatorial office) or in New Jersey. In both cases Democratic legislators took the lead (Haskins (2006), pp. 34-35). Thus, it seems worthwhile to also look at the interplay between governors and state legislatures in the process of welfare reform. This makes the setting an interesting case for studying the effects of divided government on reform-making.

Another interesting feature of welfare politics is that it is typically a very partisan topic in the US (e.g. Weissert (2000), p. 5, or Royed and Borrelli (1997), p. 543). Thus, if it was true that differing partisan positions indeed lead to gridlock under divided government, one should definitely observe this for the area of welfare politics.

2.3.2 Data

Welfare Reform. This analysis is based on a novel data set on welfare policy reform activity in US states from 1978 to 2010 that has been assembled and coded from several sources. The main dependent variable in the econometric analysis is a dummy that is equal to one if a state has conducted a welfare reform in a given year. Before the 1996 Welfare Reform at the federal level, the reform dummy is equal to one if a state has filed a welfare waiver application. The data on waivers have been obtained and cross-checked from Lieberman and Shaw (2000), Koerper (1996), and Crouse (1999). Although the dummy captures waiver applications this reflects de facto welfare policy changes since only a tiny fraction of these applications have been rejected by the federal level or withdrawn by the state level.[6] Since the 1996 Welfare Reform at the federal level, states have not submitted waivers anymore and the reform dummy is equal to one if a state has changed its welfare policy. Data on the post 1996 welfare policy changes are obtained from the Urban Institute that keeps track of all changes and maintains a large Welfare Rules Database (Urban Institute (2012)).

The baseline welfare reform dummy captures policy changes in the relevant areas of family caps, work requirements, sanctions, and time limits. A family cap rules that if a single mother on welfare conceives an additional child she does not receive additional welfare benfits for it. Work requirement rules state how many hours a welfare recipient has to work to be eligible for benefits, what the exemptions are for being ill etc. Sanctions define what happens when recipients do not comply with the rules of the system. These sanctions differ, for example, in duration and severity. Time limits state, for example, for how many years in their entire lifetime recipients are eligibile to receive benefits. For details on these rules and the coding, see the Data Appendix, in particular Appendix Table A.10. In total, the baseline reform dummy is based on changes in 14 relevant policy rules and is equal to one if a change of at least one of them occured in a particular state and year. The dummy mean over all state year observations in the sample is 0.23, i.e. in any given year about one quarter of all states reform at least some part of their welfare system. Most reforms restrict access to welfare. The ratio of contractive versus expansive reforms is more than three to one.

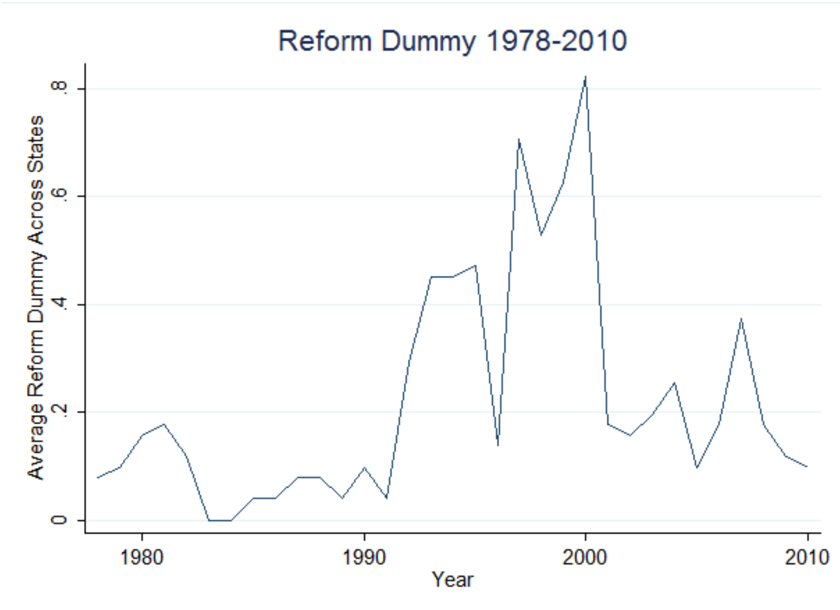
For robustness checks, alternative welfare reform dummies and count variables are constructed and used in the analysis. A narrow reform measure is based on a subset of only 8 highly relevant rules and only takes into account large changes in these rules. A broad reform measure is based on an extended set of 24 policy rules and also reflects changes in those rules. An example for such a rule in the extended set is that you need to get your children vaccinated to be eligible for welfare benefits. For details regarding data sources and coding of the welfare policy rules, see the Data Appendix. Section 2.4 gets back to the different reform measures.

The resulting data set spans the years from 1978 to 2010 and gives a comprehensive overview of welfare reform activity in US states. The distribution of welfare reforms over time is depicted in Figure 2.1. One can see that welfare reforms were especially popular in the early 1990s. Up

[6]For example, the rejection rate in Clinton's first term was 3%. Between 1978 and 1996, i.e. during the whole waiver period under consideration here, less than 8% of all waivers have ever been withdrawn, possibly by subsequent governments. For details on these numbers, see Lieberman and Shaw (2000). The approach of using application dates to measure reform is preferable over using actual implementation dates since the application date is when the political decision at the state level has been made and the implementation dates are often delayed due to interference of the Department of Health and Human Services at the federal level.

to more than 40% of states per year filed waiver applications in these years. This was the period when caseloads were high which in many cases led to the political wish to restrict access to welfare by shifting the focus of the system "from welfare to workfare". This was also the time when President Clinton announced to "end welfare as we know it". In 1996, the Welfare Reform under Clinton decentralized considerable power to shape welfare to the state level. And one can clearly see from Figure 2.1 that many states used the newly gained liberty to do so: The years from 1997 to 2000 are those in the sample with the highest number of states per year conducting welfare reforms (up to 80%). Since 2001, the share of reforming states per year has usually fluctuated around 20%. Thus, states have remained active in shaping their welfare policy rules until today.

Figure 2.1: Incidence of Reforms over Time



Divided Government. The main explanatory variable is a dummy that is equal to one if a state has a divided government in a given year. Divided government means that in at least one of the legislative chambers the majority is from a different party than the govenor. Thus, this includes so called split branch governments where the governor is confronted with majorities from the opposing party in both chambers of the legislature as well as split legislature governments where the two legislative chambers have majorities from different parties.[7] For an illustration of the different types of divided government, see Figure 2.2. Section 2.5 gets back to

[7]Nebraska has a unicameral legislature and is excluded from the econometric analysis (like Alaska and Hawaii). This is standard in the literature, see for example Lott and Kenny (1999). The analysis also excludes governments with an independent governor or split chambers (where both Democrats and Republicans have the same amount of seats).

the different types. The data on party control of state governments and legislatures have been obtained from Klarner (2003). From 1978 to 2010, more than one half of all state governments were divided. Out of these, about 60% were split brach and 40% split legislature governments.

Figure 2.2: Different Forms of State Government

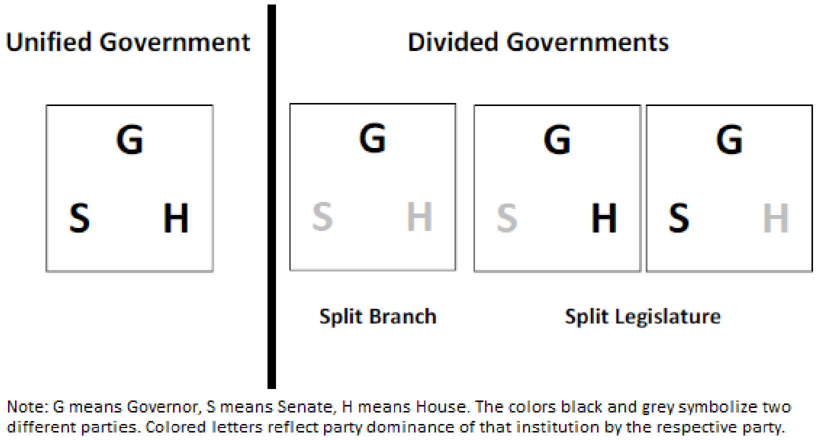
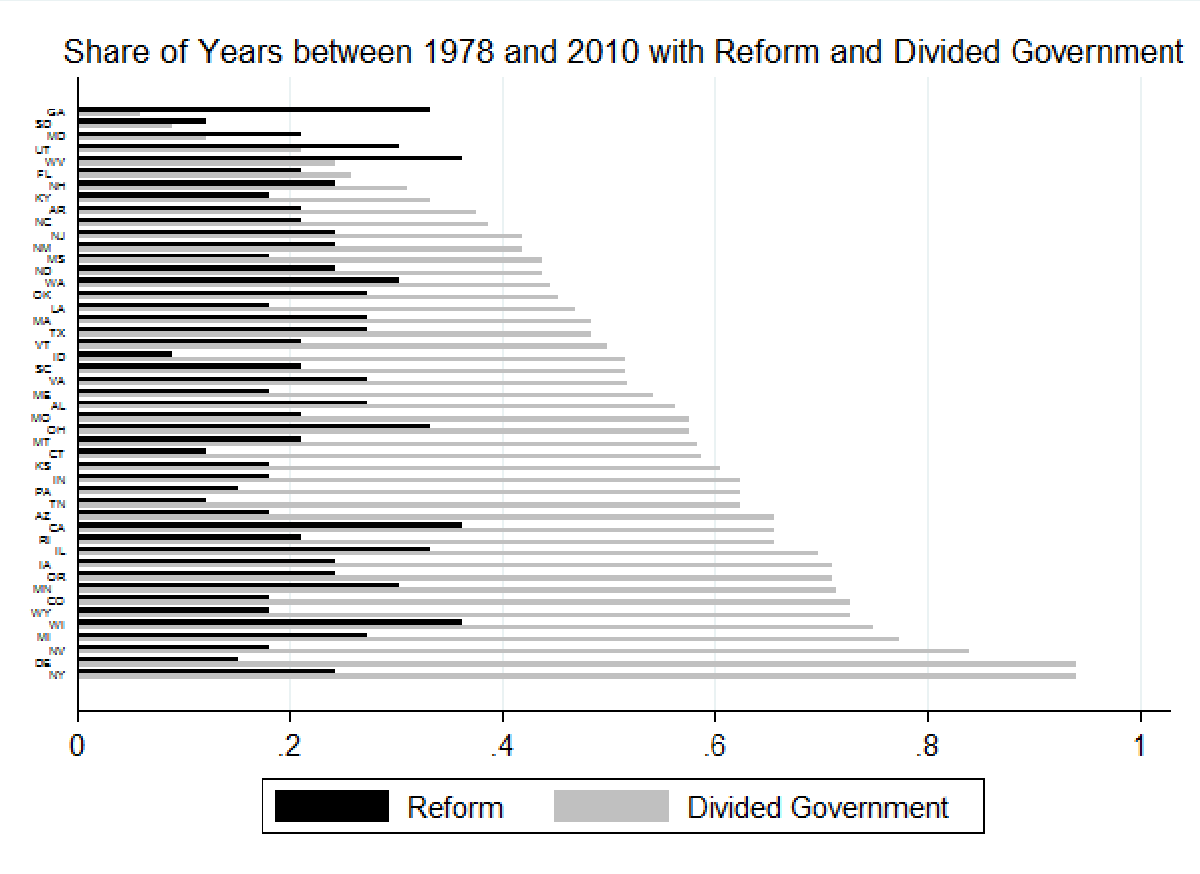


Figure 2.3 shows the cross-sectional distribution of welfare reforms and divided governments across US states. The grey bars indicate the share of years between 1978 and 2010 in which a state had a divided government. Among the states which had a divided government very often are, for example, New York and Delaware. At the opposite end, with the state government being unified almost all of the time, one finds states such as Georgia or South Dakota. Note that not a single state in the sample had either unified or divided governments for the whole time span under consideration. The black bars show the share of years between 1978 and 2010 in which a state has reformed its welfare system. These bars are on average considerably shorter than the divided government bars. Note, however, that also in terms of welfare reform years there is quite substantial variation between states. Wisonsin, for example, gets close to 40% whereas Idaho barely reaches 10%. For maps showing the distribution of reforms and divided government across states, see the Appendix.

Controls. In the econometric analysis, I control for a wide range of additional variables. Descriptive statistics of all variables are provided in Table 2.1. Means conditioned on the type of government (divided or unified) are presented in Appendix Table A.7. The demographic variables include per capita income, population size, black and latino population, population older than 65 and younger than 18. These controls are standard in US state level policy analyses. For potential relevance for welfare, I add the share of AFDC/TANF recipients (welfare caseload), the percentage of unemployed and immigrants, the deflated total state revenue per capita, unmarried birth, the maximum AFDC/TANF benefit for a family of three, and the 90th/10th ratio of

Figure 2.3: Share of Years with Reform and Divided Government



household income. Most of the demographic data are taken from the Statistical Abstract (United States Census Bureau (2011)). As political controls, I add information related to the governor (the party, if he/she can be reelected, an election year dummy), information related to the state legislature (the Demcratic seat shares in both legislative chambers, the percentage of women in the state legislature, the polarization of both chambers), information about divided government at the federal level, and ideology measures (the percentage of Democratic votes in the last presidential election and ideology measures for the state government and the state citizens taken from Berry et al. (1998)). The data have been obtained from different sources. For data sources and variables explanations, see the Data Appendix.

Table 2.1: Summary Statistics

Variable	**N**	**Mean**	**Std. Dev.**
Reform Dummy	1551	0.2302	0.4211
Broad Reform Dummy	1551	0.2585	0.4380
Broad Reform Count Variable	1551	0.5796	1.4906
Narrow Reform Dummy	1551	0.1792	0.3837
Narrow Reform Count Variable	1551	0.2469	0.6214
Reform Package (including contractive and expansive policies)	705	0.0766	0.2661
Reform Dummy Geographic Neighbors	1551	0.2241	0.2748
Reform Dummy Population Size Neighbors	1551	0.2253	0.2713
Divided Government Dummy	1474	0.5285	0.4994
Split Legislature Dummy	1474	0.1900	0.3924
Split Branch Dummy	1474	0.3385	0.4734
Divided Government with past Divided Government	698	0.3983	0.4899
Divided Government with past Unified Government	698	0.1117	0.3153
Divided Government with past Dem. Unif. Govt.	698	0.0802	0.2718
Divided Government with past Rep. Unif. Govt.	698	0.0315	0.1748
Divided Govternment via General Elections	698	0.0802	0.2718
Divided Government via Midterm Elections	698	0.0315	0.1748
Share of AFDC/TANF Recipients in Pop. (Caseload)	1551	0.0298	0.0175
% Unemployed (/1000)	1551	0.0060	0.0021
Deflated Total State Revenue per Capita (/1000)	1551	2.1730	0.6871
Unmarried Birth (per 1,000 unmarried women)	1493	28.863	9.3541
Maximum AFDC/TANF Benefit Level for a Family of 3 (/1000)	1551	0.3642	0.1373
Per Capita Income (/1000)	1551	22.688	10.299
Population (/1000000)	1551	5.5337	5.8385
% Population Black	1551	10.189	9.4377
% Population Latino	1457	6.7483	8.5792
% Population 65 or older	1551	12.430	1.8055
% Population 17 or younger	1551	26.127	2.6259
% Immigrant Population	1551	1.8831	1.9715
90th/10th Ratio of Household Income	1551	7.9765	1.3750
Governor Lame Duck (i.e. cannot be reelected)	1551	0.2650	0.4415
Gubernatorial Election	1551	0.2785	0.4484
Polarization Senate	1551	0.3488	0.1138
Polarization House	1551	0.3548	0.1086
% Women in State Legislature	1551	18.437	8.4792
% Democratic Votes in Last Presidential Election (/1000)	1551	0.0446	0.0079
Citizens Ideology (Berry et al. 1998) (/1000)	1551	0.0489	0.0154
Democratic Seat Share in Senate	1551	0.5607	0.1793
Democratic Seat Share in House	1551	0.5587	0.1716
Government Ideology (Berry et al. 1998) (/1000)	1457	0.0500	0.0242
Governor Party Dummy (1 = Democrat)	1551	0.5199	0.4962
Divided Government at Federal Level	1551	0.6061	0.4888

Notes: For details on coding, variables meanings, and data sources, see the Data Appendix. The variables "Divided Government with past Dem. Unif. Govt." and "Divided Government via General Elections" are not the same, they are highly correlated, but differ for several observations. Mean and standard deviation happen to be the same. The same is true for the variables pair "Divided Government with past Rep. Unif. Govt." and "Divided Government via Midterm Elections". See the Data Appendix for details.

2.4 Empirical Strategy and Results

2.4.1 Empirical Strategy

Panel data regressions build the main part of the analysis. The dependent variable is the welfare reform dummy. The treatment of interest is divided versus unified government which differs across states and time.

The baseline estimation equation is:

$$R_{st} = \alpha_t + \gamma_{0s} + \gamma_{1s} * t + \delta * D_{st} + X_{st} * \beta + \epsilon_{st}$$

R_{st} is a dummy that is equal to one if state s has conducted a welfare reform in year t. D_{st} is a dummy that is equal to one if state s had a divided government in year t. δ thus captures the treatment effect of interest. α_t captures year fixed effects, γ_{0s} and γ_{1s} capture state fixed effects and allow for state specific linear trends. X_{st} are relevant controls. Standard errors are clustered at the state level to take serial correlation into account (Bertrand et al. (2004)).

For simplicity, linear probability models are estimated almost throughout the paper. Linear probability models do not take into account the binary character of the outcome variable. However, estimation of fixed effects logit models yields even slightly stronger results. See Appendix Table A.3. Also on all other instances where a standard linear probability model does not seem fully adequate, alternative specifications using other estimation techniques such as Arellano-Bond or Poisson regressions are always reported in the Appendix.

Besides standard demographic controls, X_{st} includes different variables related to welfare to take potential endogeneity issues into account. One problem with identification could, for example, be that welfare state crisis is an omitted variable that may cause both divided government and welfare reform. This is why the share of welfare recipients in the population, the share of unemployed, state revenue, and other controls are included as measures of welfare state crisis. It is also known that immigration and race issues frequently come up in debates about the welfare state (Schram et al. (2003)). The analysis therefore also controls for the racial composition and immigrants in the population. For the field of welfare policy, all these socioeconomic controls can also be considered being proxies for the demand of welfare reform legislation.

Even controlling for a wide range of socioeconomic variables, several econometric concerns may remain. This is what the subsections following the main results deal with: Further potentially omitted variables such as political or ideological factors are included in the analysis.

A state's reform history and reform spillovers across states are taken into account. Another subsection deals with the measurement of welfare reform by checking alternative dummy and count dependent variables based on broader or narrower welfare policy rules sets. Potential reverse causality and other endogeneity issues are not of any concern as shown by a close elections analysis providing quasi-random variation in the type of government. For analyses regarding differential effects for different subsamples, e.g. differences depending on government history, on being a Southern state or not, or on looking before or after the federal level reform in 1996, see the Appendix.

2.4.2 Effect of Divided Government on Reform Adoption

Main Result. Table 2.2 presents the main results. The dependent variable is the reform dummy that indicicates if a state has conducted a welfare reform in a given year or not. The main explanatory variable is the divided government dummy. Specification (1) includes year fixed effects, specification (2) adds state fixed effects, specification (3) adds state specific linear time trends. In all three specifications, the effect of divided government on reform is highly significant and in the range of 4 to 6 percentage points. This means that the likelihood of observing a welfare reform is 4 to 6 percentage points higher under divided government than under unified government. Given the fact that between 1978 and 2010 the average unconditional probability of a US state to conduct a welfare reform is 23%, the effect of divided government on the probability to adopt a welfare reform amounts to more than 25% of the unconditional probability to implement a reform according to baseline specification (3). All following specifications include year fixed effects, state fixed effects, and state specific linear time trends.

Welfare State Crisis. It may be that welfare state crises are common causes of both divided government and welfare reform. Specifications (4), (5), and (6) therefore control for the share of welfare recipients, the share of unemployed in the population, and for state revenue. State revenue can be considered as a measure of fiscal crisis, results are the same when using state expenditures or state debt instead (not reported). Neither of these controls is significant, but the effect of divided government keeps its size and significance. This is also the case when adding the full range of demographic controls in specification (7). These controls include the share of immigrants, the 90th/10th percentile ratio of household income (as inequality measure), the incidence of unmarried birth (since AFDC/TANF policies sometimes aimed to reduce unmarried birth), the maximum welfare benefit for a family of three, the per capita income, the population size, the share of black or latino people, the share of people older than 64, and the share of

Table 2.2: Divided Government and Reform

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Divided Government	0.0413**	0.0552***	0.0600***	0.0598***	0.0600***	0.0602***	0.0649***
	(0.0195)	(0.0183)	(0.0204)	(0.0203)	(0.0205)	(0.0201)	(0.0229)
Share of Benefit Recipients (Caseload)				0.523			-0.201
				(2.453)			(2.481)
% Unemployed (/1000)					-0.546		-17.29
					(10.58)		(15.58)
Deflated State Revenue per cap. (/1000)						-0.0103	-0.00329
						(0.0513)	(0.0555)
Unmarried Birth							0.0126
							(0.00925)
Max. Benefit Family of 3 (/1000)							0.273
							(0.629)
Per Capita Income (/1000)							-0.0176
							(0.0175)
Population (/1000000)							0.0579
							(0.0714)
% Population Black							-0.0237
							(0.0452)
% Population Latino							-0.0260
							(0.0265)
% Population 65 or older							-0.0459
							(0.0657)
% Population 17 or younger							0.0282
							(0.0174)
% Immigrant Population							-0.00699
							(0.00787)
90th/10th Ratio of Household Income							0.00941
							(0.0142)
Year FE	YES	YES	YES	YES	YES	YES	YES
State FE	NO	YES	YES	YES	YES	YES	YES
State Specific Linear Trend	NO	NO	YES	YES	YES	YES	YES
Observations	1,474	1,474	1,474	1,474	1,474	1,474	1,343
R-squared	0.253	0.283	0.315	0.315	0.315	0.315	0.326

Notes: The dependent variable in all specifications is a reform dummy that is equal to one if one or more welfare reforms have been introduced in a given state and year. Divided Government is a dummy that is equal to one when either the majority of the state's lower legislative chamber or the majority of the state's upper legislative chamber is from another party than the governor. Unmarried birth is per 1000 unmarried women. For details regarding these or any of the demographic controls, see the Data Appendix. The demographic controls are all lagged by one year. Robust standard errors clustered at the state level are shown in parentheses. *** p<0.01, ** p<0.05 and *p<0.1.

people younger than 18. All the controls are lagged by one year since politics may need some time to react. None of the controls is significant. The effect of divided government, on the other hand, is still significant and is 6.5 percentage points large. Results are the same when taking the current values of the demographic variables or changes in the demographic variables compared to the previous year as controls (not reported).

2.4.3 Potential Other Reform Drivers

Political Factors. This subsection explores political and ideological factors and reform spillovers across states as other potentially relevant drivers of welfare reform. Table 2.3 checks the inclusion of other prominent political factors besides divided government. Specification (1) controls for lame duck governors, i.e. governors who cannot be reelected and may have different incentives. The control is not significant. Specifications (2) and (3) check if the results are affected by upcoming or just passed elections. It seems that in the year just after a gubernatoral election the reform adoption propensity is lower. Preparation of welfare reform may just take some time. The important thing to note here is that none of the gubernatorial controls affects the divided government finding. Specifications (4) and (5) include controls related to the state legislature. Specification (4) checks the effect of polarization of chambers measured as 0.5-|democratic seat share-0.5|, ranging from 0 for a fully Democratic or Republican chamber to up to 0.5 for a seat share of exactly 0.5 for each party. The significant coefficient for the polarization of the House is to be interpreted as follows: A 10 percentage points decrease in the absolute distance of the Democratic seat share from 50% (implying increasing polarization of the chamber) increases the likelihood of observing a welfare reform by 5.36 percentage points. Thus, more polarized Houses seem to be more likely to reform. In terms of interpretation, this finding fits the divided government finding. However, even when controlling for polarization, the effect of the divided government dummy itself also stays significant and keeps its size. Specification (5) controls for the share of women in the state legislature. It shows that having more women seems to reduce the likelihood of a welfare reform being adopted. This is in line with standard results in the literature.[8] The effect of divided government is not affected. The same is true for specification (6) which includes a dummy that is one for all years in which the federal government was divided. Specification (7) finally includes all political controls from before simultaneously. Again, the effect of divided government is stable and significant. Thus, even when taking into account several other political key variables, divided governments are significantly and relevantly more likely to reform the welfare system than unified governments.

Ideological Factors. A very relevant political factor in shaping welfare reform may be ideology of the state population, the state legislature, or the state governor. Table 2.4 therefore introduces several ideological controls into the analysis. Specifications (1) and (2) add the share of Democratic votes in the last presidential election and the citizen ideology measure by Berry et al. (1998). The latter measure is constructed from the ideology of state congressional delegations. See the Data Appendix for details. Neither of the two variables affects reforming or the divided government finding. Specifications (3) and (4) investigate potential effects from the partisan composition of the state legislatures. While (3) introduces the Democratic seat shares

[8]Since adopting a welfare reform typically restricts access to welfare in this data, this finding is consistent, for example, with Lott and Kenny (1999) who find that in the US extending suffrage to women came along with increases in government spending and more liberal voting by representatives, Chattopadhyay and Duflo (2004) who show that women in India implement different public good provision policies compared to men, or Funk and Gathmann (2014) revealing that women in Switzerland have stronger preferences for welfare compared to men.

Table 2.3: Divided Government, other Political Factors and Reform

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Divided Government	0.0650***	0.0650***	0.0636***	0.0597**	0.0705***	0.0649***	0.0637***
	(0.0228)	(0.0231)	(0.0228)	(0.0241)	(0.0221)	(0.0229)	(0.0234)
Governor Lame Duck	-0.00217						0.00261
	(0.0279)						(0.0285)
Year before Gubernat. Elect.		-0.00309					-0.0187
		(0.0302)					(0.0321)
Year after Gubernat. Elect.			-0.0602**				-0.0630**
			(0.0262)				(0.0295)
Polarization Senate				-0.145			-0.104
				(0.282)			(0.283)
Polarization House				0.536**			0.502*
				(0.251)			(0.260)
% Women in State Legislature					-0.0142***		-0.0138**
					(0.00510)		(0.00533)
Divided Govt. at Fed. Level						0.0377	0.0231
						(0.131)	(0.129)
Lagged Demographic Controls	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES
State FE	YES	YES	YES	YES	YES	YES	YES
State Specific Linear Trend	YES	YES	YES	YES	YES	YES	YES
Observations	1,343	1,343	1,343	1,343	1,343	1,343	1,343
R-squared	0.326	0.326	0.328	0.328	0.331	0.326	0.335

Notes: The dependent variable in all specifications is a reform dummy that is equal to one if one or more welfare reforms have been introduced in a given state and year. Divided Government is a dummy that is equal to one when either the majority of the state's lower legislative chamber or the majority of the state's upper legislative chamber is from another party than the governor. Demographic controls are lagged by one year. For details regarding any of the variables, see the Data Appendix. Robust standard errors clustered at the state level are shown in parentheses. *** $p<0.01$, ** $p<0.05$ and *$p<0.1$.

in the two chambers, (4) also interacts these seat shares with a Democratic chamber majority dummy, thus allowing partisan effects to be different depending on majority versus minority status in the chamber. None of these controls is significant, the divided government effect is stable in size and significance. Specification (5) uses the government ideology measure from Berry et al. (1998) as control, specification (6) a simple Democratic governor party dummy. Again, the divided government result is not affected. Specification (7) adds all controls from before simultaneously. This only seems to strengthen the divided government effect. Specification (8), finally, interacts the divided government dummy with the Democratic governor dummy allowing the divided government effect to be different for governors of different partisanship. Still, there do not seem to be any ideological differences. The divided government effect is still significant and reaches about 8 percentage points in size.

Ideology cannot explain welfare reform. Appendix Table A.4 reveals that robustness with respect to including ideological controls also holds when restricting attention to contractive welfare reforms only (which cover more than 75% of all welfare reforms in the data set). Also when being compared to unified Democratic and unified Republican governments separately, divided governments are more likely to reform than both (not reported). The fact that welfare reform cannot be explained by a simple story of ideology is also consistent with case studies evidence. For example, in New Jersey it was Democrats enacting a series of welfare reforms in the early 1990s that were even more drastic than Wisconsin's which had been known for strongly stressing workfare over welfare (Haskins (2006), p. 34). Another example is the federal 1996 US Welfare Reform itself which has been signed by Democrat Bill Clinton. Kansas is actually the only state where the Republicans had majorities in both chambers for the whole period of welfare reform (Weissert (2000), p. 9).

Table 2.4: Divided Government, Ideology and Reform

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Divided Government	0.0600***	0.0600***	0.0591***	0.0627***	0.0608***	0.0574***	0.0714***	0.0773*
	(0.0204)	(0.0204)	(0.0207)	(0.0203)	(0.0210)	(0.0189)	(0.0214)	(0.0425)
% Dem. Votes Presid. Elect.	-0.295						1.499	
	(2.867)						(3.373)	
Citizens Ideology (/1000)		0.0628					1.053	
		(2.010)					(2.792)	
Dem. Seats Senate			-0.0670	-0.379			-0.384	
			(0.188)	(0.262)			(0.273)	
Dem. Seats House			-0.309	-0.0293			0.0253	
			(0.258)	(0.239)			(0.255)	
Dem. Seats Senate * Dem. M.				0.168			0.203*	
				(0.104)			(0.116)	
Dem. Seats House * Dem. M.				-0.126			-0.107	
				(0.0943)			(0.103)	
Government Ideology (/1000)					-0.743		-1.425	
					(0.796)		(1.769)	
Gov. Dem. (0 = Rep.)						-0.0101	0.0334	0.0124
						(0.0235)	(0.0514)	(0.0535)
Divided Govt. * Gov. Dem.								-0.0371
								(0.0716)
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
State FE	YES	YES	YES	YES	YES	YES	YES	YES
State Specific Linear Trend	YES	YES	YES	YES	YES	YES	YES	YES
Observations	1,474	1,474	1,474	1,474	1,383	1,474	1,383	1,474
R-squared	0.315	0.315	0.316	0.318	0.327	0.315	0.331	0.315

Notes: The dependent variable in all specifications is a reform dummy that is equal to one if one or more welfare reforms have been introduced in a given state and year. Divided Government is a dummy that is equal to one when either the majority of the state's lower legislative chamber or the majority of the state's upper legislative chamber is from another party than the governor. % Democratic Votes in Last Presidential Election is divided by 1000. All seats variables refer to seat shares. Government and Citizens Ideology are both calculated from Berry et al. 1998. For details regarding these variables or the political and ideological controls, see the Data Appendix. Robust standard errors clustered at the state level are shown in parentheses. *** p<0.01, ** p<0.05 and *p<0.1.

Reform History and Spillovers. Another highly important factor potentially determining welfare reform may be learning from the own past or from others, i.e. reform could depend on yesterday's reform or be affected by policy spillovers between states. Table 2.5 explores this issue. Specifications (1) and (2) add lagged values of the reform dummy as explanatory variables. None of these is significant and the divided government finding is not affected. The same holds true when employing Arellano-Bond estimation. See Appendix Table A.5. Specifications (3) and (4) use lagged average levels of reform in geographically neighboring states as explanatory variables. Reforms in neighboring states do not seem to have an effect on a state's reform propensity. The coefficient of divided government is significant and relevant as before. This is in line with anecdotal evidence stressing how states focused on their own specific welfare programs without relying too much on the experience of neighboring states (see e.g. Liebschutz (2000), p. 18). Specifications (5) and (6) explore controlling for the lagged average level of reform in states with a similar population size. There is a positive reform adoption effect of past reforms in states with similar population size. This may suggest that states copy states with similar size when it comes to reforming. Importantly, the effect of divided government is stable in size and significance across all specifications. The same is true when adding the third or fourth lag of any of these controls (not reported).

Table 2.5: Divided Government and Reform History

	(1)	(2)	(3)	(4)	(5)	(6)
Divided Government	0.0655***	0.0653***	0.0649***	0.0647***	0.0656***	0.0653***
	(0.0235)	(0.0238)	(0.0230)	(0.0229)	(0.0227)	(0.0224)
Reform Dummy (t-1)	-0.0420	-0.0438				
	(0.0333)	(0.0347)				
Reform Dummy (t-2)		-0.0330				
		(0.0291)				
Reform Dummy Geogr. Neighbors (t-1)			-0.0843	-0.0793		
			(0.0708)	(0.0720)		
Reform Dummy Geogr. Neighbors (t-2)				0.0803		
				(0.0667)		
Reform Dummy Pop. Neighbors (t-1)					0.114*	0.111
					(0.0659)	(0.0661)
Reform Dummy Pop. Neighbors (t-2)						0.161**
						(0.0673)
Lagges Demographic Controls	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
State FE	YES	YES	YES	YES	YES	YES
State Specific Linear Trend	YES	YES	YES	YES	YES	YES
Observations	1,343	1,343	1,343	1,343	1,343	1,343
R-squared	0.327	0.327	0.327	0.328	0.327	0.331

Notes: The dependent variable in all specifications is a reform dummy that is equal to one if one or more welfare reforms have been introduced in a given state and year. Reform Dummy Geographic Neighbors is equal to the average of the reform dummy for all geographically adjacent states. Reform Dummy Population Size Neighbors is equal to the average of the reform dummy for all states with a similar population size (where all states are grouped into 10 different bands of similar population size). Divided Government is a dummy that is equal to one when either the majority of the state's lower legislative chamber or the majority of the state's upper legislative chamber is from another party than the governor. Demographic controls are lagged by one year. For details on any of the variables, see the Data Appendix. Robust standard errors clustered at the state level are shown in parentheses. *** p<0.01, ** p<0.05 and *p<0.1.

2.4.4 Measurement of Reform

The baseline reform dummy codes policy changes based on 14 highly relevant welfare policy rules. See Appendix Table A.10 for details. Nevertheless, one may argue that the reform measure should take into account all policy rule changes. Or one may argue that it should not only measure if there was a reform or not, but that it should also consider the size of the reform (e.g. number of policy rule changes) or the significance of the reform (e.g. important reforms only). This is why this subsection explores the robustness of the finding with respect to the measurement of reform.

Figure 2.4: Different Measures of Reform after 1996

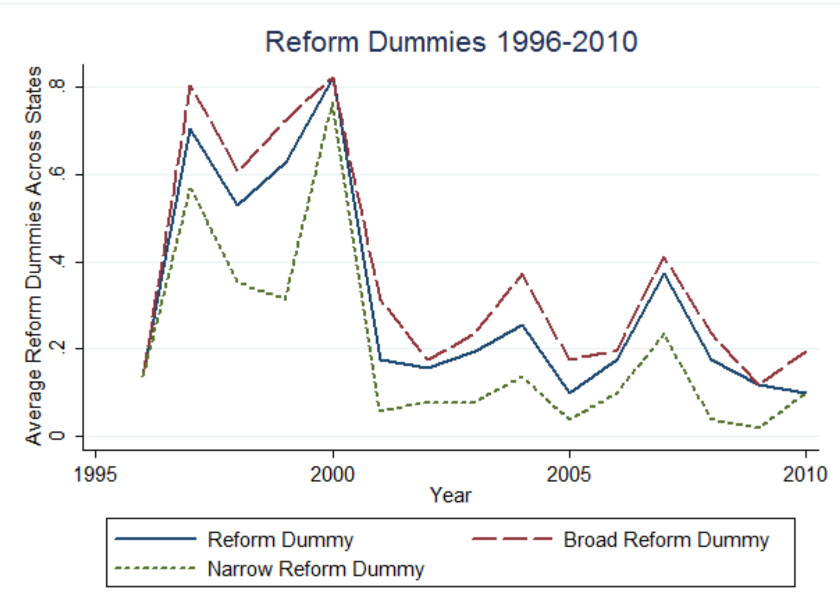


Figure 2.4 presents the development of the baseline reform dummy, of a broad reform dummy, and of a narrow reform dummy from 1996 to 2010. Note that before 1996 all measures are identical since in the waiver period the measurement of reform is unambiguous. Compared to the baseline measure, the broad dummy definition is based on 24 policy rules by considering 10 additional rules on assets exemptions and special eligibility requirements. See Appendix Table A.11 for details. In contrast, the narrow reform dummy restricts attention to a subset of highly relevant policy rules and focuses on large changes. Again, see the Data Appendix for details. As one would expect, there are more reforms when broadening the definition and less reforms when narrowing the definition. Also note that there is considerable comovement of the three different variables across time suggesting that the reform measures habe been coded in a consistent manner.

Table 2.6 explores using these alternative measures of reform as dependent variables. Specification (1) of the table replicates the baseline estimation from before using the reform dummy. Specification (2) uses the broad definition of welfare reform. The divided government effect is stable. Specification (3) relates to the argument that one should also consider the size of reforms and uses a count variable version of the broad reform dummy from specification (2). Still, the coefficient of divided government is sizable and significant. Specification (4), finally, restricts attention to important reforms only by using the narrow reform measure. Still, the divided government effect is stable in size and significance. Appendix Table A.6 shows that the results are robust when employing Poisson estimation. Thus, it is not the case that divided governments pass only minor policy adjustments, while unified governments implement large reforms. In summary, the divided government effect is very robust with respect to the measurement of the dependent variable.

Table 2.6: Divided Government and Different Measures of Reform

Dependent Variable:	**(1) Dummy**	**(2) Broad Dummy**	**(3) Broad Count Var.**	**(4) Narrow Count Var.**
Divided Government	0.0649***	0.0669***	0.129*	0.0786**
	(0.0229)	(0.0209)	(0.0724)	(0.0372)
Lagged Demographic Controls	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
State FE	YES	YES	YES	YES
State Specific Linear Trend	YES	YES	YES	YES
Observations	1,343	1,343	1,343	1,343
R-squared	0.326	0.349	0.459	0.282

Notes: The dependent variable in specification (1) is a reform dummy that is equal to one if one or more welfare reforms have been introduced in a given state and year. The dependent variables in specifications (2) to (4) are reform measures based on a broader or narrower set of welfare policy rule changes. Divided Government is a dummy that is equal to one when either the majority of the state's lower legislative chamber or the majority of the state's upper legislative chamber is from another party than the governor. Demographic controls are lagged by one year. For details regarding any of the variables, see the Data Appendix. Robust standard errors clustered at the state level are shown in parentheses. *** p<0.01, ** p<0.05 and *p<0.1.

2.4.5 Close Elections

Even when controlling for a wide range of potentially relevant variables in the analysis, some identification concerns may remain. Let me shortly outline three: First, maybe there is no causal relation between divided government and reform, but instead political competition is a relevant omitted variable causing a positive correlation between divided government and reform: States with strong political competition are more likely to implement economic reforms, but are at the same time more likely to end up with divided government. The fixed effects in the previous analyses ensure that the result cannot be driven by differences between states with strong political competition and states with weak political competition (but must be driven by within-states variation). But still, political competition may be a relevant concern. Second, reverse causality may be an issue. Assume that a gubernatorial candidate announces during her electoral campaign that she wants to reform welfare after the election and that voters do want the candidate but do not want welfare to be reformed (which may seem a priori unlikely actually). Voters may then decide to elect the candidate, but to also divide power by electing a state legislature of another partisanship than the gubernatorial candidate. Causality would run from reform intention to divided government in this case. Third, if voters know that divided governments are more (or less) likely to implement reforms, they may vote in such a way to divide (unify) government exactly when they want reforms to be implemented (and vice versa).[9]

To take into account the identification concerns just named, we would ideally need an experiment where some states are randomly assigned a divided and others a unified government. Focusing on close elections comes reasonably close to this ideal. Let us suppose that final election results are random to at least some degree. For example, rain on election day could influence the partisan composition of voters going to the polls. If one – just to fix ideas – further assumes that the state house and the state senate are both dominated by Democratic majorities, the outcome of the gubernatorial race not only determines the partisanship of the future governor, but also if government will be unified or divided. If the gubernatorial election happens to be close enough, the result of the election can be considered random, and thus also the assignment of unified versus divided government. A similar logic applies to state legislative elections and the resulting seat shares determining majority and minority status of the parties in state house and state senate.[10] Close elections – be it elections for the gubernatorial office or for the legislative chambers – can therefore provide us with quasi-random variation in the assignment of divided versus unified government.

This is why Table 2.7 looks at close elections. While specifications (1) and (2) replicate the main finding in the full sample (with and without demographic controls), specifications (3) to (6) restrict the sample to cases of close elections. The 10% sample includes all cases where ex

[9] For a general discussion of endogeneity problems arising when voters take into account the effects of institutions, see Acemoglu (2005).

[10] I use seat shares of parties in state legislative chambers to measure closeness of majorities in chambers. In principle, one could argue that using election results of individual legislators may be preferable to my approach since even an election that results in a 50% Democrats and 50% Republicans seat shares distribution (suggesting a very close race and a random election result) may in principle be perfectly foreseeable if one half of the districts are clearly Democrat and the other half are clearly Republican. But in fact seat shares should be fairly good proxies of the overall closeness of the parties' fight for the majority in a chamber. Besides, Folke and Snyder (2012) argue that seat shares may be preferable to individual vote shares since a relevant share of legislative races at the state level is uncontested which could possibly imply selection bias.

Table 2.7: Divided Government and Reform in Close Samples

	Full Sample		**10% Sample**		**5% Sample**	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Divided Government	0.0600***	0.0649***	0.0716**	0.0778**	0.0987**	0.108**
	(0.0204)	(0.0229)	(0.0297)	(0.0308)	(0.0472)	(0.0479)
Lagged Demographic Controls	NO	YES	NO	YES	NO	YES
Year FE	YES	YES	YES	YES	YES	YES
State FE	YES	YES	YES	YES	YES	YES
State Specific Trend	YES	YES	YES	YES	YES	YES
Observations	1,474	1,343	829	760	473	435
R-squared	0.315	0.326	0.354	0.381	0.435	0.467

Notes: In columns (3) to (6), the samples are restricted to observations where the election result determining whether government would be divided or unified was decided by a 5 (10) percentage points or smaller vote/seat margin. For details, see the RDD Appendix. The dependent variable in all specifications is a reform dummy that is equal to one if one or more welfare reforms have been introduced in a given state and year. Divided Government is a dummy that is equal to one when either the majority of the state's lower legislative chamber or the majority of the state's upper legislative chamber is from another party than the governor. Demographic controls are lagged by one year. For details regarding any of the variables, see the Data Appendix. Robust standard errors clustered at the state level are shown in parentheses. *** p<0.01, ** p<0.05 and *p<0.1.

post a 10 percentage points or smaller change in the governor vote margin and/or the legislative seat shares would have been sufficient to change treatment from divided to unified government or vice versa. Thus, loosely speaking, the focus is on elections where on the day before the election voters could not know already if government would be divided or unified after the election. One can see that the estimated effect of divided government on reform adoption is significant and between 7 and 8 percentage points. When focusing on even closer elections (where a 5 percentage points or smaller election result change could have induced treatment change), the effect is still significant and between 10 and 11 percentage points. This shows that political competition as omitted variable driving both reforms and divided government cannot explain the finding. If this was the case, one would expect the divided government effect to vanish when restricting attention to competitive situations. On contrary, the estimated coefficient is still significant and even gains in size compared to the full sample case. This suggests the effect of divided government on reform adoption to be even stronger in competitive situations. Overall, looking at close elections with quasi-random treatment assignment clearly illustrates the robustness of the finding of more reform-adoption under divided than under unified government.

An alternative to looking at close elections via sample restrictions like in Table 2.7 is to conduct a fully-fledged regression discontinuity design (RDD) analysis. However, the divided goverment setting poses two difficulties. First, such an RDD analysis would be non-standard in the sense of having three interdependent treatment assignment variables (election results of governor, house, and senate). This complicates the analysis quite a bit and necessitates even more data than a standard RDD. Second, the data set used in this paper is limited to a maximum of about 1,500 observations. Even for a standard RDD, this would not be much. The results of an RDD analysis reported in the Appendix do nevertheless look reassuring.

2.5 Policy Competition as Explanation

The previous sections have shown that divided governments are more likely to implement welfare reforms compared to unified governments. This does not only contradict conventional wisdom, but is also not in line with standard reasoning in the literature by models such as the one of political inaction by Howitt and Wintrobe (1995) or the veto player theory (Tsebelis (1995), Tsebelis (2002)). The latter, for example, would predict the likelihood of observing a reform to be decreasing in the number of veto players. In contrast to unified governments, in divided governments both parties are veto players. Still, we observe more reforms under divided governments. Why is this the case? – This section sheds some first light on this question.

Policy competition may be the explanation. That political competition matters for US state-level politics has been shown elsewhere for growth policies (see Besley et al. (2010)) and is also apparent for welfare politics: For example, Table 2.3 reveals that more polarized Houses reform more and Table 2.7 shows that the reform effects of divided government get larger when focusing on situations where elections have been very competitive. However, even when restricting the sample to competitively elected governments in Table 2.7, the effect of divided government on reform adoption stays highly significant. Thus, divided government as such seems to add yet another layer of competition even in competitive situations in general.

Policy Competition. The structural difference between divided and unified government is that under divided government both parties have agenda setting power. This may induce additional policy competition between them which may in the end lead to more reforms being adopted compared to unified governments where one party alone controls the policy agenda. It has been known for a long time that in US states the policy branches (executive versus legislative) engage in competition against each other. The same is true for policy struggle between the legislative chambers (e.g. Rosenthal (2009), p. 197). This policy competition across institutions may be especially strong under divided government, i.e. when these institutions are of different partisanship. For example, it may be politically too expensive for the governor to simply say no to everything that the opposition majority puts on the legislative floor.[11] Instead he or she may decide to react by a counter reform proposal. The competition may in turn lead to new

[11]Simple no-saying seems to be rather the exception in US states. See Rosenthal (2009), pp. 266, 271, 272.

ideas and finally to new policies. One example of such a policy competition story under divided government could be that the opposition majority leader uses her agenda setting power in the legislature to pass innovative bills in order to qualify as future policymaker and governor in the eyes of the voters.[12] This also fits the fact that more than one third of all US governors have been members of legislatures before (Gray and Hanson (2008), p. 194).

Case studies on US welfare reforms indeed suggest that policy competition is relevant during the reform process, especially under divided government. Let me illustrate this by giving short examples from four states which all had divided governments when they passed significant welfare reforms: In New York, Republican governor Pataki struggled long and hard with the Democratic state legislature, especially the assembly speaker and the senate president. Together with the governor, the latter two were the "key actors" on welfare reform (Liebschutz (2000), pp. 19 and 59). In Florida, Republican senate president Jennings and Democratic governor Chiles were the key leaders to create the Florida WAGES program (Liebschutz (2000), pp. 19 and 38). In New Jersey, the Democratic assembly majority leader Bryant was even the main mover of welfare reform (Haskins (2006), p. 34). And in Wisonsin, the Democratic legislature "attempt[ed] to outdo the governor [Republican reformer Tommy Thompson] in welfare reform" (Wiseman (1996), p. 532). Although the governor was the leader on welfare reform, it was Democratic majorities in the state legislature passing the reforms and sometimes suggesting even more radical reforms than Thompson (Haskins (2006), p. 35 and Liebschutz (2000), p. 109). These examples clearly emphasize the role of competition among parties and their key representatives under divided government. Another fact that may speak in favor of a "positive competition" argument is that reforms implemented under divided government are actually statistically not more likely to be reverted later than reforms passed by unified governments.

The best example of reform competition under divided government is probably the passage of the US Welfare Reform at the federal level in 1996. Although Clinton had campaigned on welfare reform in 1992 and the Democrats controlled both chambers of Congress in 1993 and 1994, not much happened during the first two years of the Clinton administration (Haskins (2006), p. 37). Democrats were themselves divided on the issue of welfare reform, could not agree on a unified proposal, and focused on other agenda items such as health instead (Haskins (2006), p. 39). However, in the November 1994 midterm elections Republicans got a majority in both Senate and House and immediately made use of this newly gained agenda setting power to pass a radical welfare reform bill in January 1995 (Haskins (2006), pp. 20, 192). Although

[12] For a similar idea, see Mayhew (2005), p. 105. For an illustration of the importance of legislative majority leaders, see Rosenthal (2009), pp. 226 and 236.

Clinton vetoed the bill in December 1995 (Haskins (2006), p. 253), he was challenged and had to politically react. Republican Speaker Newt Gingrich was a key figure in pushing for welfare reform and negotiating it with Clinton. After yet another veto in January 1996 (Haskins (2006), p. 266), Clinton finally signed a version of the bill in August 1996. The most important welfare reform since the New Deal had been passed under divided government.

If policy competition is indeed the explanation for the divided government effect on reform adoption, one may in fact expect effects of differential size of different types of divided government. For an overview of these different types, consider again Figure 2.2. Under unified government, all agenda setting power is with one party. Competition should be at its minimum. Under split branch governments, one party has the executive agenda setting power, the other party has full control of the legislative agenda. This may induce policy competition across the two branches. Under split legislature governments, however, competition may be even more intense since policy struggle may now not only take place between branches, but also within the legislature (between the legislative chambers). Following this interpretation of the competition argument, one would expect unified governments to reform the least, split branch governments to reform more, and split legislature governments to reform even more again. This is a prediction one can empirically assess.

Specification (1) of Table 2.8 does no longer use the divided government dummy as explanatory variable, but instead the two subtypes split branch and split legislature government. The results show that split branch governments are about 5 percentage points and split legislature governments about 9 percentage points more likely to reform than unified governments. Although the difference between the two coefficients is not significant, the result is still indicative. Thus, the empirical evidence is at least consistent with the policy competition theory. Next to the policy competition theory, let us also shortly empirically investigate three other potential explanations: a blaming theory, a pet reform theory, and a signaling theory.

Blaming. A blaming theory could go as follows: If a governor implements a reform under unified government, she will be held fully accountable by voters in case the reform happens to fail. If, in contrast, she passes the reform under divided government, she may be able to blame the opposition party for the failed reform.[13] This way the governor may be able to shift part of the responsibility for a failed reform onto the legislature dominated by the opposition party. This may induce governors to be more willing to suggest reforms under divided government. Under divided government, they are insured against reform failures and do not need to fear to

[13]See Andersen et al. (2010) on evidence that governors are only punished by voters for late budgets if their party also has a majority in the state legislature. See Alt and Lowry (1994) for similar evidence on fiscal policy.

Table 2.8: Explaining the Relationship between Divided Government and Reform

Dependent Variable:	**Reform Dummy**		**Reform Package**	**Reform Dummy per Govt.**	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Sample Restriction:				**All Gov. Any Elect.**	**Non-LD Gov. Midterm Elect.**
Split Legislature Dummy	0.0908**				
	(0.0346)				
Split Branch Dummy	0.0529**				
	(0.0255)				
Divided Government		0.0744***	0.0223		
		(0.0265)	(0.0368)		
Divided Government * LD		-0.0378			
		(0.0488)			
LD		0.0174			
		(0.0342)			
Δ Seats Gov's Senate Faction				-0.238	-1.202
				(0.180)	(0.915)
Δ Seats Gov's House Faction				0.0959	1.066
				(0.185)	(0.880)
Lagged Demographic Controls	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES
State FE	YES	YES	YES	YES	YES
State Specific Linear Trend	YES	YES	YES	YES	YES
Observations	1,343	1,343	665	667	229
R-squared	0.326	0.326	0.227	0.486	0.675

Notes: The dependent variable in all specifications (1) and (2) is a reform dummy that is equal to one if one or more welfare reforms have been introduced in a given state and year. The dependent variable in specification (3) is a dummy that is equal to one if there was a reform in a given state and year that included expansive and contractive policy rules changes at the same time. The dependent variable in specifications (4) and (5) is a dummy that is equal to one if one or more welfare reforms have been introduced by a government in a given state. Divided Government is a dummy that is equal to one when either the majority of the state's lower legislative chamber or the majority of the state's upper legislative chamber is from another party than the governor. Split Branch is equal to one if the governor in a state is confronted with majorities of the opposing party in both legislative chambers. Split Legislature is equal to one if the majorities in the two legislative chambers in a state are from opposing parties. LD (lame duck) is equal to one if the governor cannot be reelected. Demographic controls are lagged by one year. For details regarding any of the variables, see the Data Appendix. Robust standard errors clustered at the state level are shown in parentheses. *** $p<0.01$, ** $p<0.05$ and *$p<0.1$.

be held fully accountable in terms of votes in the next election.

How does this blaming theory fare empirically? – Note that it should be easier for governors to shift the blame for a failed reform onto the legislature when both legislative chambers are controlled by the other party. We should therefore expect to see more reforms under split branch governments than under split legislature governments. As seen in Table 2.8 before, this is not the case in the data. Another empirical test of the blaming theory exploits potential differences between reelectable and non-reelectable governors: The above outlined blaming should only be relevant for governors who are still afraid of being held accountable by voters for failed policy reforms. Lame duck governors who cannot be reelected anyway should just engage in the reforms they consider best without thinking about how divided governments may insure them against failures. If the blaming theory is correct, we should therefore not expect to see more reforms under divided government in the case of lame duck governors. This is what specification (2) of Table 2.8 checks by introducing a lame duck control and also an interaction with the divided government dummy. One can see that neither new variable is significant, but the divided government effect is stable. The divided government effect thus seems not to be differential for lame duck governors.[14] From both empirical checks we can therefore conclude that it is unlikely that a blaming theory can explain the finding. Another issue with this theory is that it could also go the other way round: If a governor is unwilling to reform, she may be more willing not to implement a reform under divided government since she can always (falsely) blame the other party for the reform gridlock.

Pet Reforms. The pet reform theory suggests that each politician has "pet reforms" he or she really wants to implement. Under divided government, more ideologically diverse politicians have a relevant say in policymaking than under unified government. And more ideologically diverse politicians have more ideologically diverse pet reforms. This may lead to more reforms being adopted under divided government. One particular channel could be that under divided government Republicans and Democrats just pack together a rightist change of one policy and a leftist change of another policy into one reform such that each party gets what it wants. "Dirty compromising" could be going on. If this was true, we would expect to see reform packages including at least one leftist policy change and at least one rightist policy change at the same time to be more likely under divided than under unified government. But, in contrast to this theory, specification (3) of Table 2.8 reveals that divided government does not increase the likelihood of observing such packages.

Signaling. Lastly, this leaves us with the signaling theory. This time, interpret divided government as a result of frustrated voters punishing an incumbent governor for policy disappointments by taking away the governor's majority in the legislature in midterm elections (Alesina and Rosenthal (1995) and Folke and Snyder (2012)). This may signal to the governor that voters are not happy with her performance in office and that the likelihood may be large that voters will not reelect her in the next election. In this desperate situation, the governor may then engage in risky reforming in the hope of being lucky with it and consequently being reelected. If we interpret divided government as a signal to the incumbent governor in this way, we may expect to see more policy reforms under divided than under unified government.[15]

[14] Bernecker and Gathmann (2013) focuses on governors and their reputational concerns during the US Welfare Reform in greater detail. This analysis includes further factors such as the governor's age or previously hold offices (potentially measuring quality or ambition for higher office), but still finds the positive divided government effect.

[15] A similar theory would be that if the governor loses midterm she may ideologically move towards the opposing party in order to get closer to the median voter's bliss point by implementing reforms the opposing party supports. This would also result in more reforms after midterm losses of the governor.

What about empirical evidence consistent with this theory? – Since this theory again relies on governors mainly caring about their reelection prospects, one would again expect to see differences for lame ducks versus reelectable governors. But specification (2) of Table 2.8 shows that this is not the case. As another check one can directly investigate the reform effect of governors losing votes midterm. Specification (4) of Table 2.8 uses the seat share changes in the senate and house factions of the governor party in the last election as explanatory variables. According to the signaling theory, one should expect a negative relation with reforming: Drops in seat shares should lead to more reforms. However, the estimated coefficients have opposing signs and both are not significant. Specification (5) restricts the sample to reelectable governors who just faced midterm elections, i.e. it focuses on governors who fit exactly the story outlined above. Still, there are no effects. Thus, also the signaling theory is likely not the explanation of the effect of divided government on reform adoption.

Finally, one could also hypothesize that under divided government parties of different partisanship share the political responsibility and that politicians of different partisanship jointly can signal reform necessity to voters in a more credible way than politicians of just one party. The idea would be a bit along the lines of Cukierman and Tommasi (1998) who show that a leftist politician may be more able to credibly signal to voters the necessity of a rightist policy compared to a rightist politician. The better signaling ability of divided governments may then result in more reforms being passed by them compared to unified governments. However, this story is unlikely to be applicable in the case of the US Welfare Reform since there was a broad consensus also among the population that welfare politics needed to change (Mead (2004), p. 8).

Summary. In summary, one can at least say that the policy competition theory is consistent with reform case studies and the empirical evidence. The data seem to contradict other potential explanations. However, to get a deeper understanding of what is going on in state-level reform politics and what exactly causes divided government to make reforms more likely, one would need richer data than used for this paper. One first step could be to analyze bill sponsorship and voting on reforms at the individual legislator level. Such data is unfortunately unavailable for most of the time span under consideration and such an analysis is beyond the scope of this paper.

2.6 Conclusion

Conventional wisdom suggests that under divided government political parties block each other resulting in a lack of economic reforms. This paper systematically tests this view by analyzing novel data on welfare policy reforms at the US state level between 1978 and 2010. Panel data estimates show that the probability to implement welfare reforms is in fact between 5 and 10 percentage points *higher* for states with divided government as opposed to states with unified governments. This effect amounts to 20 to 50% in size of the unconditional probability to implement a welfare reform and is robust with respect to the inclusion of a wide range of control variables. The effect also keeps its significance and size when using alternative measures of reform, when employing different estimation techniques, and when focusing on close elections providing quasi-random variation in the type of government. One potential conclusion is that in fact voters do not have to worry too much about reform gridlock from electing a divided government. This may be part of the reason why many voters repeatedly prefer divided over unified government not only in elections, but also when being asked about it in polls (Alesina and Rosenthal (1995), p. 44).

The finding that divided governments are more likely to reform can potentially be explained by stronger policy competition between governor, senate, and house compared to the case of unified government. The underlying idea is that the opposition party can use the agenda setting power it is endowed with under divided government to confront the governor as key policy-maker with policy proposals. Empirical evidence is consistent with the competition theory and inconsistent with other possible explanations of the finding. Case studies evidence also supports the competition theory. But given the limitations of the data set analyzed in this paper further research is definitely needed to shed more light on the explanation of the finding.

Given that more than one half of all US state governments are divided, my results are certainly very relevant in the US context. But, even if a prominent one, it is only one example. In Western democracies in general, unified governments seem to be rather the exception than the rule (Fiorina (1996), p. 111). In France, for instance, the term "cohabitation" is used to describe a very similar phenomenon that occurs when the president faces a majority of an opposing party in parliament and therefore has to appoint a prime minister of this opposing party. Also in many parliamentary democracies different party control of different institutions is often argued to result in blockades which supposedly make reforming impossible. Take the example of Germany where very often the second chamber (consisting of members of state governments) has a different bloc majority than the first chamber (the parliament electing the federal government). Since most important laws need a majority in both chambers, legislative deadlock can possibly arise. Given my result, one may have to rethink common deadlock claims made with respect to divided government also for these and other countries. But further research is needed until any conclusions about reform-making in different political contexts can be drawn.

Chapter 3

Trial and Error? Reelection Concerns and Policy Experimentation during the US Welfare Reform[1]

3.1 Introduction

When President Clinton signed the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) in August of 1996, it marked the most dramatic reform of US welfare policy since the New Deal. The law abolished the entitlement to cash assistance for poor families, the "Aid to Families with Dependent Children"(AFDC). The newly established "Temporary Aid to Needy Families"(TANF) defined welfare benefits as temporary assistance rather than an entitlement. PRWORA further decentralized the authority to design welfare policies to the individual states (see, for example, Grogger and Karoly (2005), Moffitt (2008)). After August 1996, state policy-makers were free to choose their own eligibility criteria for benefit receipt, which requirements recipients had to fulfill while on welfare and what sanctions would

[1]This chapter has been co-authored with Christina Gathmann. For helpful comments and discussions, we thank Toke Aidt, Manuel Arellano, Henning Bohn, Pierre Boyer, Matz Dahlberg, Peter Fredriksson, Vincenzo Galasso, Mario Jametti, Dylan Minor, Espen Moen, Jörg Oechssler, Torsten Persson, Per Pettersson-Lidbom, Federico Revelli, Christian Riis, Elisabeth Schulte-Runne, Christoph Vanberg, and Galina Zudenkova. We further thank seminar participants at universities in Freiburg, Heidelberg, Konstanz, Norwegian Business School, the Silvaplana Workshop on Political Economy, the Conference on Evaluation of Political Reforms in Mannheim, the European Public Choice Society Meeting, the CESifo Conference on Public Sector Economics, the International Institute of Public Finance Annual Meeting, and the European Economic Association Meeting. We are very grateful to Greg Shaw, Dave Andersen, and Paul Ehmann for sharing some of their data, to Shelly Arsenault for providing supplemental material on state waivers and the administrators of the Welfare Rules Datebase (WRD) at the Urban Institute for answering our detailed questions. Jan Scherer, Michael Hellwig, and Robert Aue provided outstanding research assistance. An earlier version of this chapter has been circulated in the University of Mannheim Economics Department Working Paper Series (Bernecker and Gathmann (2013)).

be imposed if recipients did not comply with these requirements.

A vast literature has analyzed the consequences of the 1996 welfare reform for caseloads, employment, earnings, poverty and many other outcomes (see, for example, Blank (2002), Grogger and Karoly (2005), Moffitt (2001), and Moffitt (2002b) for excellent surveys). Little is known, however, how state governments actually decided on their welfare policies after decentralization (but see Soss et al. (2001)).[2] Many welfare rules had not been implemented before and most states had little experience which welfare rules would work in their particular circumstances. If little is known about the effectiveness of certain policy rules, there are potentially large gains from learning via experimentation. Yet, how much did states take advantage of this opportunity to experiment both before and after PRWORA? More importantly, what determined how much states experimented, and whether they adapted their policies after observing the experiment's consequences?

In this paper, we set out to provide a first answer to these questions. Our basic argument is that reelection concerns play an important role for the decision to experiment and possibly reverse welfare rules later on. The idea is that experimentation with policy rules reveals information not only to politicians but also to voters. If politicians differ in their ability to identify appropriate policies (and voters do not observe a politician's competence), a politician may try to manipulate voters' beliefs about her competence. A politician might then experiment strategically to signal competence to voters. If the experiment is successful, it improves the politician's reputation among voters and hence her reelection chances. The same logic makes politicians too hesitant to reverse a failed policy experiment because voters would interpret this as evidence of low ability.

To capture this idea, we (verbally) outline a simple political agency model of reputation concerns in which both voters and politicians are uncertain about the best welfare policy (building on Majumdar and Mukand (2004)). During her term in office an incumbent can decide whether to stick to the known status quo policy or experiment with a new, uncertain policy. Policy-makers and voters observe the policy chosen and outcome midterm. In the next period, the politician can decide whether to stick with the policy or (if she experimented) revert to the status quo. The outcome of this policy choice is observed prior to the next election.

Based on the model, we derive three empirical predictions when politicians experiment and possibly reverse an experiment later. The decision to experiment depends on three key parameters: the quality of the politician, the potential gain from the experiment for society, and

[2] In contrast, several detailed studies track the political process leading up to the federal reform in 1996 (Reintsma (2007), Weaver (2000), or Haskins (2006)).

the initial reputation of the politician. First, high-quality politicians will experiment more the higher the gains from experimentation as this increases overall output. Low-quality politicians in contrast might experiment less in this situation because the potential reputational gains from successful experimentation are low. Second, experimentation further depends on the initial reputation of the incumbent. The higher her initial reputation, the more an incumbent has to lose if the policy experiment does not work out. We would therefore expect a negative correlation between an incumbent's initial reputation and policy experimentation. Third, the quality of the politician and the potential gain from the experiment also influence the decision to reverse a policy experiment later on. A low-quality politician is less willing to reverse an experiment if the potential gains from the experiment are large. The reason is that successful experimentation yields a large reputational boost and thus increases the incumbent's reelection chances.

To test our predictions in the context of the US welfare reform, we collect a new dataset on welfare policy rules under AFDC and TANF (i.e. before and after PRWORA). Our analysis thus spans both the period of welfare waivers states adopted under AFDC as well as the period following the federal reform in 1996. We complement our panel of policy rules with additional information on the political process, characteristics of state governors and measures of the socioeconomic conditions in each state.

Our results suggest that reelection concerns matter both for the decision to experiment or reverse a policy experiment later on. Consistent with the theory, we find that high-quality governors are more likely to experiment when the potential gain from experimentation are high. The effect is sizeable: A one standard deviation increase in governor quality and the gain from experimentation raises the propensity to experiment between 37% and 49%. A higher initial reputation of the politician has the opposite effect: It reduces the propensity to experiment. Again, the effect is sizeable: A one standard deviation increase in initial reputation reduces experimentation by about 27%. We also find evidence for the third prediction: Low-quality governors are hesitant to reverse a policy when the potential gains from sticking with the experiment are high. A one standard deviation increase in the potential gain reduces the propensity to reverse an experiment by 36%. We also show that governors that cannot be reelected (and hence are less concerned about their reputation among voters) behave systematically differently than governors that can be reelected.

A number of robustness checks bolster our findings that reelection concerns indeed matter for the decision to experiment. First, our main analysis abstracts from ideological considerations. However, voters and politicians might care about the particular policy rules implemented in their state. During the US welfare reform, for instance, Republicans were typically more in favor of workfare rather than welfare. They might therefore be more eager to experiment with stricter rules. Surprisingly, we find that neither governor nor citizen ideology influence the decision to experiment conditional on state and year fixed effects. More importantly, controlling for governor party, government ideology or citizen ideology has no effect on our results regarding reelection concerns.

Second, our analysis focuses on the governor as the relevant decision-maker abstracting from legislative pressures and initiatives. Yet, controlling for the party composition and polarization of the legislature has again no effect on our results.

Third, our main analysis abstracts from spillover effects across states. We therefore create several measures of experimentation in neighboring states where neighbors are defined in terms of a shared border or, alternatively, similar population size. Conditional on state and year fixed effects, we find no systematic relationship between the propensity to experiment and the degree of experimentation in neighboring states over the past three years. Not surprisingly, including controls for potential spillover effects between states has therefore little effect on our estimates.

A fourth concern could be that we do abstract from the local demand for welfare services. Incentives or pressures to experiment might vary depending on the current economic condition or the social structure in a state. To check this, we include a number of time-varying state characteristics to our baseline specification. Yet, including the unemployment rates, income per capita, wage inequality, and immigrant share, size, age and racial composition of the population has little effect on our main results.

In sum, experimentation with welfare waivers during the AFDC program and welfare rules during the TANF program both support the idea that electoral incentives (and reputation-building by the governor) are important for understanding policy innovations and the incentive to reverse policy experiments.

The paper is structured as follows. The next section discusses the related literature while Section 3.3 provides background information about the US welfare reform. In Section 3.4, we introduce theoretical thoughts on policy experimentation and reversals and derive the empirical predictions. We describe the empirical strategy, data sources, and descriptive statistics in Section 3.5. Section 3.6 presents the results and tests the robustness of our evidence. Finally, Section 3.7 concludes.

3.2 Related Literature

Our article contributes to a number of different literatures. First, our starting point is closely related to the idea of laboratory federalism. In his survey, Oates (1999) summarizes this idea as follows: "In a setting of imperfect information with learning-by-doing, there are potential gains from experimentation with a variety of policies for addressing social and economic problems. And a federal system may offer some real opportunities for encouraging such experimentation and thereby promoting 'technical progress' in the public sector." (Oates (1999), p. 1135).

A sizeable literature in political science studies innovation and diffusion of policies such as the adoption of state lotteries or labor market regulation (e.g. Walker (1969), Cnudde and McCrone (1969), Allen and Clark (1981), Berry and Berry (1990), Berry and Berry (1992), or Boehmke and Skinner (2012); see Karch (2007) and Berry and Berry (2007) for good surveys).

Our analysis differs from prior studies on policy innovation along several dimensions. First, we provide the first analysis of policy experimentation during the US welfare reform, considered one of the major social reforms in the US since the New Deal.[3] While there are several accounts of the period leading up to the federal welfare reform in individual states (Francis (1998), Tweedie (2000) or Weissert (2000), for example), our study provides systematic quantitative evidence where there are rich historical case studies.[4] Furthermore, we focus on the systematic influence of reelection concerns for the decision to experiment with waivers and policy rules. Second, in contrast to other studies, we also analyze the decision to experiment as well as the decision to later reverse a policy. Third, earlier studies are typically not tied to a theoretical model. In contrast, we rely on political agency considerations showing how reelection concerns matter for policy experimentation and reversals. We use the testable predictions to motivate our empirical analysis.

Our theoretical setup is related to a large literature on reputational concerns following the seminal work of Holmström (Holmström (1982), Holmström (1999)). An early application to the political arena is the political agency model (Rogoff (1990)). Closer to our setting 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), or Willems (2013)). As in our setting, politicians have an incentive to manipulate voters' beliefs

[3]One exception is Lieberman and Shaw (2000) who analyze the choice of welfare rules as a function of local conditions in the context of the US welfare reform.

[4]There is one study analyzing why certain states chose restrictive policies under the TANF program (Soss et al. (2001)). The authors analyze a variety of factors including electoral competition, ideology and the racial composition of the welfare population.

about the politician's ability to enhance their reelection chances. The focus in our paper is on bringing some predictions of these models to the data. Our analysis is the first to test the implication of reputational concerns for incentives to innovate and learn in the political arena.

An alternative explanation for policy experimentation relies on the politician's ideology. Politicians can boost their reputation, for example, if they propose and implement policies that are on the other side of their ideological spectrum (see Cukierman and Tommasi (1998), extended by Moen and Riis (2010)). In our model, there is no conflict which policy is best per se. There is uncertainty whether a policy is appropriate in the current state of the world. The empirical predictions we derive are thus different from those implied by a reputational model with ideological considerations. Our empirical results clearly show that ideology cannot explain the patterns of policy experimentation we observe during the US welfare reform.

Our study is also closely related to the literature on the question why certain 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)), a war of attrition between politicians (Alesina and Drazen (1991)) or vested interests benefitting from the status quo (Coate and Morris (1999)). Instead, we focus on the reputational costs of policy reversals as one explanation why policies might persist. The most important difference is that we test our predictions empirically in the context of the US welfare reform.

Further, our analysis contributes to a small empirical literature studying the role of electoral concerns for taxes and expenditures (see Besley and Case (1995b), Besley and Case (1995a) or List and Sturm (2006), for examples).

3.3 The 1996 Welfare Reform in the United States

3.3.1 AFDC program and Welfare Waivers

Aid to Families with Dependent Children, created in the Social Security Act of 1935, was an entitlement program that provided financial assistance to needy children lacking parental care or support (see Grogger and Karoly (2005), Moffitt (2008), or Bitler and Hoynes (2010)). The program was jointly administered by the federal and state governments. Federal rules determined most of the eligibility criteria and provisions under AFDC. Generally speaking, single-parent families with dependent children who satisfied a few other criteria were eligible for aid.[5]

[5] 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.

Compared to the post-1996 period, states had relatively little room under AFDC to design and shape welfare policy. States set benefit levels which varied substantially across states, from $119 in Mississippi to $720 in Alaska in 1995 for a family of three with no other income. In addition, states could implement additional eligibility rules. For example, several states passed "fit parent" or "suitable home" provisions to limit payments to families with unsatisfactory behavior.

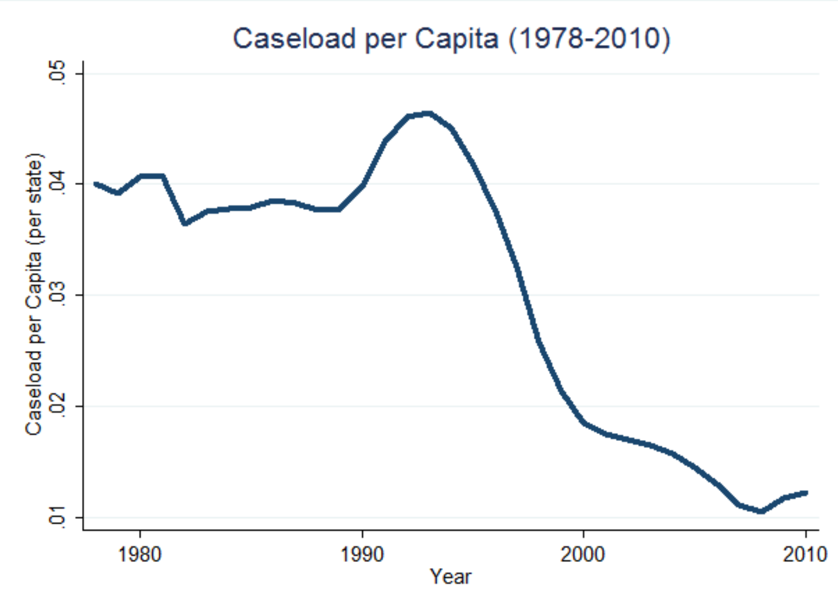
Several attempts were made to reform the AFDC program in subsequent decades. An important initiative to reduce caseloads, facilitate work participation among recipients and advance other objectives of the AFDC program was taken by state governors through welfare waivers. Since 1962, Section 1115 of the Social Security Act granted authority to the Secretary of Health and Human Services to waive the federal rules and regulations governing AFDC. Under this provision, states could petition the U.S. Department of Health and Human Services (DHHS) to implement experimental, pilot, or demonstration projects.[6]

To apply for a waiver, states submitted detailed requests which rules and program elements were modified and what new rules and regulations were implemented. The proposed waiver provisions were then reviewed, and sometimes modified, 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 requested by DHHS were incorporated into the application. In our empirical analysis, we will use the number of waivers a state applied for - irrespective of whether the waiver was approved, amended, withdrawn or denied - as a measure of policy experimentation during the AFDC period.

Welfare waivers were rare until the late 1980s when AFDC caseloads approached almost 4 millions families or around 11 million recipients nationwide (see Figure 3.1 for the evolution of per capita caseloads averaged across states). Waivers became a popular policy instrument to reduce state caseloads when federal aid for AFDC was cut during the Reagan years and funds became tight during the 1989-1992 recession. Between 1987 and 1992, 15 waiver applications in 14 states were approved during the Reagan administration and another 15 applications from 12 states were approved during 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 representing 43 states and the District of Columbia. In total, all but five states received approval for one or more waivers.

[6]These experiments (which were often implemented in a few local areas) were required to be cost-neutral and to include a rigorous evaluation of the project.

Figure 3.1: Evolution of Caseload per Capita 1978-2010



A prominent role in pushing welfare reform on the political agenda was played by state governors. Wisconsin's Republican governor Tommy Thompson, a leading state figure of welfare, is a good example. He made welfare reform a top priority in his campaign for governor as early as 1986 (Mead (2004), Kaplan (2000)). When elected in 1987, Thompson immediately created a task committee to reform the existing AFDC system. Between 1987 and 1996, the Thompson administration applied for three waivers in 1988, 1992 and 1993. In 1988, Wisconsin was the first state which conditioned a household's receipt of benefits on the school attendance of its teen children.

Under Thompson's leadership and using his line item veto power, the new Wisconsin Works, better known as W-2, was implemented in 1993. With its focus on labor market participation with stricter work requirements and harsh sanctions for noncompliance, the W-2 program became a role model for the federal TANF 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) or Winston (2002) for descriptive case studies).

As the example of Wisconsin shows, most waivers implemented multiple changes to welfare rules simultaneously (see Crouse (1999) and Koerper (1996) for detailed descriptions of the state waivers). Most popular were experiments to impose stricter work requirements (to increase work participation among recipients), impose time limits on benefit receipt, 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).[7]

3.3.2 The Introduction of "Temporary Aid to Needy Families"

The passage of the Personal Responsibility and Work Opportunity Reconciliation Act (PROWRA) in August of 1996 abolished AFDC and replaced it with the "Temporary Assistance to Needy Families"(TANF) program.[8] PRWORA removed most federal requirements on how to run the program and instead gave state governments a lot of freedom to create their own welfare programs, however generous or restrictive. During the year following the passage of PRWORA, states began to replace their existing AFDC programs with state TANF programs. Implementation dates ranged from September 1996 for Massachusetts, Michigan, and Vermont to January 1998 for California. Waiver provisions remained in place even after the passage of PRWORA until the waiver expired or was replaced by new statewide TANF rules. Because states had to decide on many new policy dimensions, policy rules in the states changed even after the adoption of TANF at the state level and continue to do so until today.

Reducing caseloads (and welfare dependence more generally) as well as boosting work participation were dominant operational goals of the TANF program. To achieve these goals, PRWORA radically departed from the AFDC program along many dimensions. For example, it introduced the idea of time limits to benefit receipt, i.e. the idea that state support was a temporary measure rather than a permanent source of income.[9] Another important new element of welfare reform was the focus on work requirements. The work requirements in the new legislation were much stronger than in previous law and changed the orientation from education and training to work per se. The law further allowed states to impose sanctions on recipients for failure to comply with the work requirements, sanctions which were much stronger than in past law and which were rigorously enforced.

Federal funding for the TANF program was consolidated into a single block grant.[10]

[7] The data appendix provides more detailed information on these new rules first implemented during the AFDC waiver period and later in state TANF programs.

[8] State governors again played a prominent role here. The National Governors' Association (NGA) was influential in lobbying for welfare reform at the federal level (see Weaver (2000)). Fourteen governors testified in welfare hearings of the 104th Congress, compared to only three state representatives (Winston (2002)).

[9] The federal government sets a lifetime time limit for benefit receipt of 60 months. States that wish to offer longer time limits (or no time limit at all) need to use state funds to support recipients for more than 60 months. States could however also decide to implement shorter lifetime time limits on benefit receipts. We return to a more detailed discussion of specific welfare policy rules in the data section.

[10] Under the AFDC program, in contrast, a matching grant by the federal government co-financed state and local contributions. The costs of additional caseloads in a state were therefore shared by the state and the federal government. After PRWORA, the contribution of the federal government is limited by the block grant. Any additional funding, e.g. rising caseloads because of bad economic conditions or less restrictive eligibility criteria, have to be borne by the state.

PRWORA imposed financial penalties on the states for not complying with the federal mandates, such as required work participation rates of recipients or caseload sizes. These penalties took the form of percentage reductions in the block grant allocation for each type of violation.[11]

States therefore had strong incentives to reduce caseloads and boost work participation rates among welfare recipients as failure to meet the required work participation rates or caseload reduction can exact big financial penalties. While all states share the focus on work first and a reduction of welfare dependence, a lot of heterogeneity exists in tone, structure, spending, and operations. Some states try to meet TANF's performance requirements by engaging a large and growing share of their cash assistance caseloads in work activities. Others reduce caseloads, either by moving beneficiaries into jobs or providing them with alternatives to cash assistance.[12]

3.4 A Model of Reputational Concerns

To study the decision to experiment with (and possibly reverse) welfare policies in the US states, we outline a simple model of reputation concerns and highlight its empirical predictions. Given the dominant role played by state governors both during the waiver period and the TANF period, we focus on a single political decision-maker who is concerned both about welfare and reelection (following Majumdar and Mukand (2004)).

3.4.1 Basic Setup

Suppose for simplicity that a high-quality politician is perfectly informed about the right policy. She will then experiment with the new policy if it is best to do so. She will not experiment if the status quo policy is better. A high-quality politician will never reverse a policy experiment even if the performance of the policy is disappointing at first. A low-quality politician in contrast, will try to mimic the high-quality politician in order to increase her reelection chances. As a consequence, a low-quality politician might experiment to make voters' believe she is a high-quality politician. A low-quality politician will also be hesitant to reverse a policy experiment because of its high reputational costs (as it signals to the voter that the politician is indeed low-quality).

[11]Each state's block grant was funded at the annual spending level corresponding to fiscal years 1992 to 1995. States were also required to contribute substantially to other program components that were replaced by TANF, for example subsidized child care.

[12]The incentives seem to bear fruit: For example, caseloads decreased by more than 56% nationally between 1994 and 2000. Employment rates for single mothers with young children grew from 68 to 78% between 1995 and 1999. See Blank (2002), pp. 1115 and 1116.

The incumbent politician can choose between keeping the current policy or implementing a new policy. The return from the status quo policy is public information, but both voters and politicians are uncertain about the returns to the new policy. The returns from the innovative policy may be higher or lower than the status quo: Social welfare will be higher if the innovative policy is the appropriate policy for the state's particular circumstances.

Politicians differ in their quality which is private information. High-quality politicians are better able than others to identify and implement policies that best fit the current state of the world. Voters do not know whether the observed outcome of the new policy is the result of a high-quality politician (making the right policy choice) or of a low-quality politician whose policy choice turns out to be successful. After the first period, the incumbent can decide to revert back to the old status quo policy (at no cost) or stay with the implemented policy experiment. Because politicians care about their reelection as well as about welfare, they might experiment with new policies for strategic reasons, i.e. to make voters believe that they are of high quality.

Suppose first the politician only cares about social welfare (i.e. the outcome of the chosen policy). In this first best case, the decision to experiment with a new policy and the decision to reverse a previously implemented policy depend on the expected net welfare gain alone. Hence, the decision to experiment will depend on the costs of implementing or reversing the policy (e.g. administrative costs) and the expected welfare gain from experimentation or reversal. For policy reversals, the decision will in addition depend on the observed outcome of the policy experiment.

Majumdar and Mukand (Majumdar and Mukand (2004)) demonstrate that there is a unique Bayes Nash equilibrium in the formalized version of this setting characterized by: (a) two cutoff rules (whether a politician experiments or not and whether she reverts back to the status quo policy or persists with the experiment), (b) a set of voters' beliefs about the politician's quality after observing the policy choice and outcome, and (c) the reelection probability for the incumbent politician given voters' beliefs.

In equilibrium, politicians sort themselves into three groups: those that do not experiment, those that do experiment but later revert to the status quo, and those that do experiment and stick to the policy experiment. Uniqueness of the sorting equilibrium is achieved by three specific features of the model: First, there are only two types of politicians (high-quality and low-quality) and the high-quality politician is perfectly informed about the appropriate policy. Second, the analysis studies electoral incentives only within a single electoral cycle. And third, the cutoff rule that defines the decision to revert or persist with a policy experiment is independent of the cutoff rule that defines whether to experiment or not.[13]

3.4.2 The Role of Reelection Concerns

Suppose now that politicians also care about their reelection, for example, because they derive some ego rents from being in office. Voters want to reelect high quality politicians who are better able to identify and implement the policy that is most appropriate for the current state of the world. Voters observe the chosen policy and a noisy signal of the outcome of policy decisions. They use this information to update their beliefs about the politician's ability. The politician then has an incentive to strategically manipulate voters' beliefs in order to increase her reelection chances.

Take as example a high-quality politician who gets a perfect signal about the state of the world (and hence, knows what is the appropriate policy to choose). A high-quality politician will therefore innovate if the new policy is appropriate (the policy that produces the largest expected net benefit). Also, she will always persist with a policy experiment (even if the observed outcome fails to be realized) because she knows that the policy experiment is the right one for the current state of the world.

A low-quality politician instead might try to manipulate voters' beliefs by mimicking the high-quality politician. She might choose to experiment even when the expected net benefit of the new policy are low because the reputational gains of experimenting are large if it is successful (i.e. the appropriate answer to the current, but uncertain state of the world). In that case, voters will update their beliefs that the politician is of high quality. However, we also might get too little experimentation from low-quality politicians. If the expected net gains from experimentation are high, experimentation (among all politicians) becomes more likely. In that case, the reputational gains from experimentation are low, reducing the low quality politician's incentive to implement a policy experiment.

Strategic considerations also influence the decision to revert back to the status quo (when the outcome of a policy experiment has been publicly observed). Reputation concerns might induce a low-quality politician to inefficiently persist with an experiment even if the policy is not appropriate for the current state of the world. The reason is that a policy reversal would signal to the voter that the politician is of low quality (because a high-quality politician never reverses a policy). Thus, if reputation concerns are sufficiently important (relative to welfare

[13]The model can be generalized to multiple types as Majumdar and Mukand show in an earlier working paper version. This more general model yields similar predictions than the baseline setup discussed here.

considerations), the politician may experiment either too much or too little and she might also be hesitant to reverse a policy.

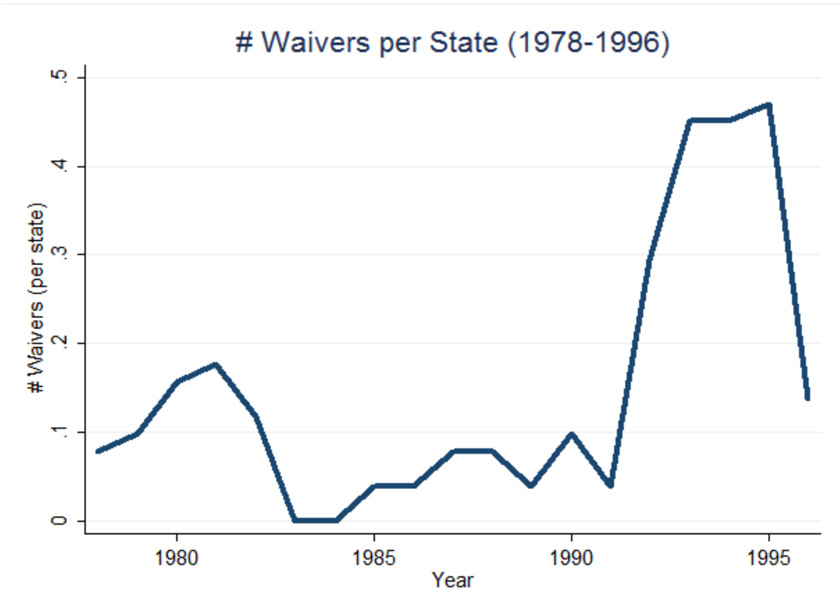
The propensity to experiment and reverse a policy experiment also depends on the politician's initial reputation. The higher her initial reputation, the fewer are the potential gains from experimenting in order to manipulate voter beliefs. Hence, a high initial reputation should make reckless experimentation less likely. Similarly, a high initial reputation makes it less likely that a policy is reversed because the reputational costs of policy reversals are high.

3.5 Data and Empirical Strategy

3.5.1 Data Sources

We collect a comprehensive new dataset of state-level welfare policy rules, political conditions, and socio-economic characteristics. Here, we summarize our key sources and variables. The Data Appendix provides a more detailed description of all data sources and the construction of our key variables.

Figure 3.2: Evolution of Waivers per State 1978-1996



Policy Experimentation and Reversals. To measure policy experimentation during the AFDC period (prior to 1996), we use the waivers states applied for at the Secretary of Health and Human Services from Koerper (1996), Crouse (1999), and Lieberman and Shaw (2000). We include 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. We use information on all waiver applications irrespective of whether the state planned to implement it only in specific counties or the whole state. Altogether we have information on 195 waivers states applied for between 1978 and 1996. Although state governments could apply for waivers since 1962, waivers became a relevant policy instrument only in the late 1970s. Figure 3.2 shows that waiver activity was low during most of the 1980s, but became a very popular tool for states in the early 1990s.

Table 3.1: Distribution of Welfare Waivers (1978-1996)

Year	# Waiver Applicants	State Codes of Waiver Applicants	Average # Waivers States Applied for
1978	4	CA, UT, VT, WI	1.0
1979	5	Co, MA, MI, NY, WI	1.2
1980	8	CA, CO, CT, FL, GA, IL, NJ, WV	1.5
1981	9	FL, IL, KY, MA, NJ, OH, TX, WI, WV	1.1
1982	6	NY, OH, OK, TX, WV, WY	1.0
1983	0	-	0.0
1984	0	-	0.0
1985	2	CA, IL	1.0
1986	2	CA, MN	1.0
1987	4	IA, MS, WA, WI	1.8
1988	4	AL, CA, NY, OH	1.0
1989	2	CA, MD	1.0
1990	5	GA, MN, OK, UT, WI	1.0
1991	2	OH, VA	1.0
1992	15	CA, GA, IL, MD, MI, MO, NJ, OK, OR, SC, UT, VA, VT, WI, WY	1.5
1993	23	AR, CA, CO, CT, FL, GA, HI, IA, IL, MA, MO, MS, ND, NH, NM, OH, OR, SD, TX, VA, WA, WI, WY	1.3
1994	23	AR, AZ, CA, GA, IL, IN, KS, MA, MD, ME, MI, MO, MT, ND, NM, NY, OH, OK, OR, PA, SC, VA, WI	1.4
1995	24	CT, DE, FL, GA, HI, IL, IN, LA, MA, ME, MO, MS, NC, NH, OH, OK, OR, SC, TX, UT, VA, WA, WI, WV	1.5
1996	7	CA, IA, MD, MI, MN, TN, UT	1.6

Sources: Lieberman and Shaw (2000), Koerper (1996), Crouse (1999).

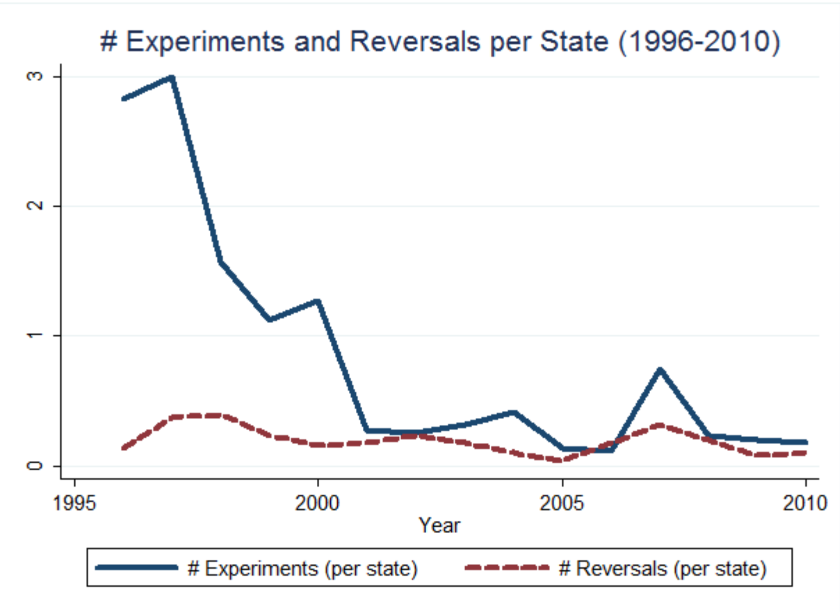
Our measure of policy experimentation prior to 1996 is then a simple count variable of the number of waivers a state applied for in a year. If a state did not apply for a waiver, the count variable is set to zero. Table 3.1 shows the distribution of waivers across states. The last column shows the average number of waivers states applied for in each year. Most states only apply for one waiver within a year. In twenty cases, a state applied for two waivers within the same year, in seven cases for three waivers and in two cases for four waivers. The maximum number was filed by California and Illinois which applied for six waivers each in 1994 and 1992 respectively.

To measure policy experimentation and policy reversals after 1996, we collect the welfare policy rules in each state and year from the Welfare Rules Database of the Urban Institute (Urban Institute (2012)). The Welfare Rules Database provides the most comprehensive and up-to-date information on welfare rules in all fifty states over the period from 1996 to 2010. The database contains literally hundreds of rules on eligibility, benefit calculation, and many other aspects of welfare reform.

In our analysis, we focus on a subset of policy rules that were at the center of the public and academic debate surrounding the US welfare reform: (1) Whether the state has a family cap in place. Under a family cap, benefits received by the household do not increase if a child was conceived and born while the parent is on welfare. (2) The rules governing time limits for benefit receipt. (3) What work requirements adults have to fulfill in order to remain eligible for benefit receipts. (4) The sanctions imposed by a state if a household does not fulfill the work requirements. Table B.1 in the Data Appendix provides more detailed explanations about the individual policy rules.[14]

Based on these rules, we define a policy experiment in a year when a state changes a policy rule for the first time. Similarly, a policy reversal is coded as one if a state reverses a rule that had been adopted in 1996 or later and still was in place in the previous year. Both variables are zero if no policy experimentation or reversal took place in a state in that year. Note that we focus on the first experiment and possible reversal later on in a state to avoid defining each incremental change as an experiment or reversal.

Figure 3.3: Evolution of Experiments and Reversals per State 1996-2010



Take, for example, family caps which require that benefit payments do not increase if the parent conceives an additional child while receiving welfare benefits. Family caps did not exist under the AFDC program, though a few states experimented with it during the waiver period. Between 1996 and 2010, twenty-four states adopted a family cap, eight of these twenty-four

[14]To test the sensitivity of our results to the selection of policy rules, we construct alternative measures using a broader set of policy rules. Specifically, we add rules about demographic and financial eligibility requirements, about schooling, health and immunization requirements for continued eligibility. In addition, we also construct measures that take into account the direction of the policy reform. In particular, we define as an experiment if welfare rules become more restrictive after the TANF reform; a reversal is then if the state relaxes the tightening of the rules later on. Our results are similar using these alternative measures of policy experimentation and reversal.

states later abolished the family cap previously adopted. We code a policy experiment if a state first adopts a family cap after 1996 and we code a policy reversal if a state (that had adopted a family cap until 1996 or later) abolishes it in any year between 1996 and 2010. We code policy reversals for other policy rules in a similar way: the first change in a policy rule is coded as an experiment, abolishing that change is coded as a policy reversal.[15] Figure 3.3 shows the evolution of experimentation and policy reversals during the TANF period. We observe 109 policy reversals (in 14% of the state-year observations). In most cases (N=85), states reverse a single policy rule in a given year (in 20 cases, states reversed two rules and in 4 cases, three policy rules in the same year).

Electoral Incentives. Our model of reputational concerns emphasizes two characteristics of the governor: the quality of the governor and her initial reputation. Both concepts are difficult to measure as they are ultimately unobservable. To proxy for governor quality, we use the years a governor has spent in all political offices before being elected as a governor (see the Data Appendix for details). Cumulative years in political offices will be a good indicator of government quality if elections select the most able candidate or candidates learn about policy-making from spending time in political office. While the number of years a governor has spent in the current office is in principle observable to voters, most voters do not know about the total political experience a governor has accumulated over her career.[16]

The second important characteristic of the governor in the model is her initial reputation after being elected. We use the vote margin in the past gubernatorial race as an indicator of

[15] Note that we can measure the adoption and abolishing of specific policy rules unambiguously only for the years since 1996 (when covered by the Welfare Rules Database of the Urban Institute). Changes in policy rules during the waiver period are therefore not captured because the information contained in the waiver application cannot be merged to the Welfare Rules Database in a consistent manner. This feature may result in two biases. If a state both adopted and abolished a family cap prior to 1996, we will not capture this in our measure of policy reversal (resulting in a downward bias in the number of policy reversals). Anecdotal evidence however suggests that policy reversals were rare before 1996 as many waivers only expired after the federal TANF reform in 1996. 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 that state adopts a family cap after 1996, we code it as an experiment in the post-1996 period (resulting in an upward bias in the number of policy experiments in the TANF period). Our post-1996 experimentation measure is not biased if changes in policy rules under waivers are still in place in 1996.

[16] Formal educational qualifications might be an alternative, albeit observable proxy as well (used in Galasso and Nannicini (2011) and Gagliarducci and Nannicini (2013) for Italian politicians or Kotakorpi and Poutvaara (2011) for Finnish members of parliament). In our analysis, educational qualifications were never found to be a statistically or economically relevant determinant of policy innovations. One explanation could be that our measures of formal education (no college degree, BA or equivalent or Master degree and other postgraduate degrees) might not contain enough variation to identify an effect as most of the governors in our sample are highly skilled: only 6% do not have a college degree or more, while almost 50% have a master or other postgraduate degree. The average local politician in Italy has a high-school degree only. Even in the Finnish parliament, only about 20-30% of the members of parliament have a master degree or more. Hence, formal educational credentials could be more a prerequisite for a governor rather than a signal of quality.

Table 3.2: Summary Statistics

	Waiver Period (1978-1996)			TANF Period (1996-2010)		
Variable	N	Mean	Std. Dev.	N	Mean	Std. Dev.
# Policy Experiments (Waiver Applications)	969	.2012384	.5765372			
Indicator for Waiver Experimentation	969	.1496388	.3569012			
# Post 1996 Experiments				714	.8137255	1.550562
Indicator for post 1996 Experimentation				714	.3431373	.4750897
# Post 1996 Reversals				714	.1918768	.4950143
Indicator for post 1996 Reversals				714	.1526611	.3599126
Governor Quality (Years in Political Office)	903	11.43189	9.096402	664	12.04518	9.759293
Low Governor Quality	903	.4983389	.5002743	664	.4623494	.4989563
Governor Age	935	55.15187	9.585323	694	59.00576	8.527857
Governor Lame Duck	950	.2378947	.4260186	650	.3169231	.4656355
Governor Past Vote Margin	913	8.367755	7.063581	674	12.45428	11.17812
Governor Party	950	.5871744	.4886193	700	.4364286	.4923173
Governor Party Ideology (Berry et al. 1998)	950	-.0185552	.2790226	700	.0593503	.3188724
Citizens Ideology (Berry et al. 1998)	950	46.96159	15.2925	600	51.29706	15.99161
Democratic Presidential Vote Share	850	41.53648	7.089873	703	47.08796	9.035134
Divided Government	931	.5488722	.4978732	679	.5478645	.4980706
Democratic Seat Share Upper House	931	.601858	.1859281	686	.5116806	.1621021
Democratic Seat Share Lower House	931	.5934898	.1800033	686	.5193672	.1535433
Polarization Upper House	931	.1703237	.1261511	686	.131253	.0957144
Polarization Lower House	931	.1638331	.1195023	686	.1252652	.0907574
% AFDC/TANF Recipients in Population (Caseload)	950	.0404008	.0144649	713	.0187619	.0139151
Unemployment Rate	950	6.465789	2.110039	711	5.289733	1.88728
Per Capita Income (/1000)	969	15.71918	5.310862	714	32.96318	7.456562
Population (/1000)	969	4.782394	5.19861	714	5.715069	6.360649
% Black Population	950	9.399228	9.157036	709	10.85607	10.71709
% Population 65+	950	11.94104	2.27207	709	12.72292	1.789727
% Immigrant Population	951	1.761939	2.215588	714	2.260454	1.718433
Gini Coefficient (Family Income)	950	.3575397	.0246252	400	.4014018	.0220994
Av. # Experiments in Adjacent States	950	.2219035	.4076696			
Av. # Experiments in States with similar Population	969	.2034744	.3502401			

Notes: For details on the welfare experimentation and reversal measures, see the data appendix. The quality of a governor is measured by years of political experience, see the data appendix for details. The low quality governor indicator is equal to one if the number of years of political experience is below the median value. Governor Lame Duck is equal to one if the governor cannot run for reelection. The past vote margin 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 absolute deviations of the democratic seat share from 50%. Adjacent states are geographically adjacent states, states with similar population refers to states with similar population size (where all states are divided into ten bands of similar population size).

initial reputation (collected from List and Sturm (2006) and Leip (2012)). Governors which voters believe to be capable are likely to be elected with a larger majority than governors that just barely reach a majority of votes. One issue with such a measure of initial reputation is that states with stronger inter-party competition (e.g. swing states) might have tight elections even independent of the governor's initial reputation. To control for this alternative influence, we check whether our results hold if we control for the competitiveness of elections other than the governor's race such as state legislative elections.

Our model also suggests that governors should behave differently in the absence of reelection concerns. We therefore collect 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. Information on gubernatorial elections and term limits are taken from List and Sturm (2006), Council of State Governments (2012), and Leip (2012).[17]

Whether policy experiments (and reversals) are undertaken also depends on the gain of an experiment. In our setting, states experiment to learn about the effects of alternative welfare policy rules. An important motivation for state governors to apply for a waiver was to encourage work and reduce caseloads which had been rising in the late 1980s and early 1990s. The incentive to reduce caseloads and increase work participation rates has been even stronger after the federal welfare reform in 1996. The main reason is that states only receive a block grant from the federal government and hence have to cover any additional costs if caseloads rise or the state does not comply with federally mandated work participation rates. As a proxy for the potential gain from experimentation, we use the lagged caseload per capita.[18]

Other Political Variables and Demographics. To control for other political influences on the decisions to experiment and revert, we collect a number of additional political data. Most importantly, we have several variables measuring the ideology of the governor: the party of the governor and a measure of governor ideology (based on Berry et al. (1998)). The latter variable takes into account differences in ideology between different governors of the same party. It varies between 0 and 100 where larger values represent a more liberal attitude.

To control for the influence of the legislature (both lower and upper chamber), we collect data on the party composition and political polarization (Klarner (2003)). Polarization is measured

[17]Even governors who face a binding term limit might have some interest in their reputation, for example, because they plan to run for another office at the federal level. For our analysis, we do not require the assumption that lame ducks have zero weight on reputational concerns (and hence only care about welfare). All we need is that lame ducks care somewhat less for their reputation among voters in their respective state than governors who can and want to be reelected in the next election.

[18]Our results are robust with respect to using alternative measures for the net gain of experimentation such as absolute instead of per capita caseloads, welfare expenditures, or welfare expenditures per capita.

as the Democratic seat share in the lower (upper) chamber calculated as deviation from 50%. We further code whether the government in the state is politically divided between Democrats and Republicans, i.e. whether the party of the ruling governor is different from the majority in the state parliament (also from Klarner (2003)).

Finally, policy innovations and reversals might also be influenced by the particular conditions in a state. We collect a number of variables to proxy for the demand side: personal income per capita, the size of the population, the share of the Black population in a state and the age structure in a state (all from United States Census Bureau (2011)). We add measures of income inequality (authors' calculation from the Current Population Survey (Center for Economic and Policy Research (2012))) and the size of the immigrant population (U. S. Department of Homeland Security (2011)). To control for citizen ideology (and possibly demand for redistribution) more directly, we use the vote share for the Democratic candidate in the last presidential election (United States Census Bureau (2011)). Table 3.2 contains summary statistics of all variables used in our analysis separately for the waiver period (1978-1996) and the TANF period (1996-2010).

3.5.2 Empirical Strategy

To test the predictions of our reputation model for policy experimentation and policy reversals during the US welfare reform, we use a simple difference-in-differences framework similar to Besley and Case (1995a) or List and Sturm (2006). To analyze the determinants of policy experimentation, we estimate variants of the following model:

$$\begin{aligned} Exp_{st} = \beta_1 Quality_{st} + \gamma_1 Gain_{st} + \delta_1 Quality_{st} * Gain_{st} + \\ + \lambda_t + \theta_s + \mu_1' X_{st} + \epsilon_{st} \end{aligned} \tag{3.1}$$

where θ_s are state fixed effects, λ_t denote year fixed effects, and X_{st} represents additional time-varying state or governor characteristics. Our first prediction suggests that higher-quality governors are more likely to experiment if the gains from experimentation are high, i.e. $\delta_1 > 0$. A low quality politician in contrast is less likely to experiment when the potential gains are high because the reputational gains from experimentation are low.

The model also predicts that a high initial reputation right after the election makes experimentation less likely when reputational concerns matter. The reason is that a high initial reputation makes it more likely that the governor is reelected than a governor with low reputation. Therefore, a politician with high initial reputation has no need to engage in risky policy experiments to boost his reputation. A politician with low reputation in contrast, would be more likely to experiment in the hope of grabbing large reputational gains. To test this prediction, we estimate the following model:

$$Exp_{st} = \beta_2 Initial\_Reputation_{st} + \lambda_t + \theta_s + \mu_2' X_{st} + \epsilon_{st} \tag{3.2}$$

where the variables are defined as before. We expect policy experimentation to decrease with the incumbent's reputation, i.e. that $\beta_2 < 0$.

Our third prediction relates to policy reversals. We are particularly interested in low quality politicians where strategic considerations are likely to be important. In the model, low-quality politicians are unwilling to revert a policy because then voters learn that the politician is of low quality. The reputational gains of sticking with the innovative policy are especially large when the net gain from the experiment are high and the policy happens to be successful in the second period. Hence, we estimate variants of the following model:

$$\begin{aligned} Reverse_{st} = \beta_3 Low\_Quality_{st} + \gamma_3 Gain_{st} + \delta_3 Low\_Quality_{st} * Gain_{st} + \\ + \lambda_t + \theta_s + \mu_3' X_{st} + \epsilon_{st} \end{aligned} \tag{3.3}$$

where $Low\_Quality_{st}$ is a binary indicator if the quality of the governor is below the median. All other variables are defined as before. We expect that $\delta_3 < 0$. We now turn to a discussion of the empirical results.

3.6 Empirical Results

3.6.1 Governor Quality, Initial Reputation and Policy Experimentation

Quality and Experimentation in AFDC Period. To test our first prediction, we regress the number of waiver applications in the AFDC period on the quality of a governor (years in prior political offices), the potential net gain (caseload) and the interaction between the two.[19] Table 3.3 shows the results. Specifications (1) and (4) include year fixed effects, all other columns additionally include state fixed effects. In columns (3) and (6), we also include the age of the governor to adjust for the fact that political experience might rise mechanically with age due to

[19] The number of waiver applications is a count variable truncated at zero and with integer increments. Thus, a Poisson model might be more appropriate. We find similar results when using a Poisson model instead.

life-cycle considerations. In line with the model, we find that more able governors are more likely to experiment with a waiver when the potential gains from experimentation are high (significant positive effect of the interaction term). Interestingly, the age of the governor is not related to the propensity to experiment.

Table 3.3: Electoral Incentives and Policy Experimentation in Waiver Period (1978-1996)

	All Governors			Differential Effect for Lameducks		
	(1)	(2)	(3)	(4)	(5)	(6)
Quality * Caseload in State	0.635***	0.491*	0.511**	0.733***	0.583**	0.586**
	(0.189)	(0.245)	(0.249)	(0.206)	(0.267)	(0.283)
Quality	-0.0207***	-0.0170*	-0.0181*	-0.0243***	-0.0192**	-0.0195*
	(0.00710)	(0.00892)	(0.00918)	(0.00770)	(0.00934)	(0.0104)
Caseload in State	-1.313	-5.647	-5.817	-1.416	-5.438	-5.402
	(1.874)	(5.156)	(5.153)	(1.855)	(5.188)	(5.186)
Governor Age			0.000871			0.000508
			(0.00156)			(0.00176)
LD * Quality * Caseload in State				-0.436**	-0.399	-0.316
				(0.217)	(0.258)	(0.349)
LD * Quality				0.0163**	0.00992	0.00537
				(0.00791)	(0.00861)	(0.0132)
LD * Caseload in State				-0.550	3.235*	1.326
				(1.014)	(1.916)	(4.038)
LD * Governor Age						0.00166
						(0.00291)
Year Fixed Effects	YES	YES	YES	YES	YES	YES
State Fixed Effects	NO	YES	YES	NO	YES	YES
Observations	846	846	846	846	846	846
R-squared	0.195	0.284	0.284	0.200	0.287	0.287

Notes: The dependent variable in all specifications is the number of policy experiments (waiver applications) in a given state and year. Specifications (1) and (4) include year fixed effects, all other specifications additionally include state fixed effetcs. The quality of a governor is measured by years of political experience, see the data appendix for details. Caseload refers to the percentage of AFDC/TANF recipients in the population, lagged by one year. The LD (lame duck) dummy is equal to one if the governor cannot be reelected. Standard errors clustered at the state level are shown in parentheses. *** $p<0.01$, ** $p<0.05$ and *$p<0.1$.

Our model also predicts that reputational concerns should matter more for governors who can be reelected. The right-hand side of Table 3.3 tests whether governors facing a binding term limit (lame duck) respond differently than governors for which reputational concerns are (more) important. And indeed, for lame ducks we find a negative interaction effect at least partially cancelling out the positive effect found before (although the negative "correction term" for lame ducks is not statistically significant in specifications (5) and (6)).

The coefficient on the interaction term is statistically significant and economically sizeable in all specifications. A one standard deviation in the gain from experimentation and quality of the governor increases the propensity to experiment by about 37%.[20] Note that high caseloads

[20]Using the coefficients in column (6), a one standard deviation of the interaction term quality*caseload (0.43) multiplied with the estimated coefficient (0.586) yields 0.25. A one standard deviation in governor quality (9.09 years) multiplied with the estimated effect (-0.0195) yields an effect of -0.18. Hence, both effects together imply

in the past per se are not systematically related to policy experimentation. The coefficient is negative suggesting that states with higher past caseloads actually experiment less and is never statistically significant. High-quality governors are per se less likely to experiment. One explanation for this pattern could be that governors with longer careers in politics are more likely to belong to the political establishment which could make them more hesitant to experimentation and change in general.

Table 3.4: Electoral Incentives and Policy Experimentation in Post-TANF Period (1996-2010)

	All Governors			Differential Effect for Lameducks		
	(1)	(2)	(3)	(4)	(5)	(6)
Quality * Caseload in State	0.799**	1.226**	1.228**	0.739	1.550*	1.671*
	(0.375)	(0.580)	(0.582)	(0.599)	(0.875)	(0.911)
Quality	-0.0124*	-0.0112	-0.0124	-0.0117	-0.0173	-0.0223
	(0.00684)	(0.0105)	(0.0111)	(0.00887)	(0.0132)	(0.0140)
Caseload in State	-13.69**	-16.72	-17.31	-14.14**	-15.39	-17.96
	(5.605)	(14.63)	(14.94)	(6.926)	(15.63)	(16.39)
Governor Age			0.00352			0.00614
			(0.00628)			(0.00684)
LD * Quality * Caseload in State				0.208	-0.344	-0.620
				(0.638)	(0.824)	(1.036)
LD * Quality				-0.00276	0.0129	0.0214
				(0.00862)	(0.0123)	(0.0205)
LD * Caseload in State				-1.112	-6.011	-0.631
				(8.851)	(10.22)	(15.98)
LD * Governor Age						-0.00297
						(0.00518)
Year Fixed Effects	YES	YES	YES	YES	YES	YES
State Fixed Effects	NO	YES	YES	NO	YES	YES
Observations	666	666	666	621	621	621
R-squared	0.383	0.406	0.407	0.379	0.403	0.403

Notes: The dependent variable in all specifications is the number of policy experiments (policy rule changes) in a given state and year. Specifications (1) and (4) include year fixed effects, all other specifications additionally include state fixed effects. The quality of a governor is measured by years of political experience, see the data appendix for details. Caseload refers to the percentage of AFDC/TANF recipients in the population, lagged by one year. The LD (lame duck) dummy is equal to one if the governor cannot be reelected. Standard errors clustered at the state level are shown in parentheses. *** $p<0.01$, ** $p<0.05$ and *$p<0.1$.

Quality and Experimentation in TANF Period. Do we find the same influences on policy experiments during the TANF period? Table 3.4 estimates the same model as before where we now use the number of policy experiments after the US federal welfare reform in 1996. We find very similar results: high-quality governors are significantly more likely to experiment in the TANF period when the gains from experimentation are high. The right-hand side again shows that this relationship again seems to be stronger for governors who can be reelected (although the difference to lame duck governors is not statistically significant for the TANF

an effect of 0.07 more experimentation. Relative to the mean share of experiments in a given year (0.2), we get an average increase of 37.30%.

period as can be seen from the non-significant interaction effects for lame duck governors). The coefficients imply that policy experimentation during the TANF period increases by about 49% with experimentation gain and governor quality. The effect is therefore even somewhat larger than during the waiver period.[21]

Table 3.5: Initial Reputation and Policy Experimentation

	All Governors			Differential Effect for Lameducks		
	(1)	(2)	(3)	(4)	(5)	(6)
Past Governor Vote Margin	-0.00680**	-0.00477*	-0.00447	-0.00869***	-0.00759**	-0.00743**
	(0.00305)	(0.00276)	(0.00277)	(0.00307)	(0.00332)	(0.00366)
Governor Age			0.000652			0.000293
			(0.00145)			(0.00144)
LD * Past Governor Vote Margin				0.00491*	0.00744**	0.00816*
				(0.00263)	(0.00281)	(0.00481)
LD * Governor Age						-0.000212
						(0.00121)
Year Fixed Effects	YES	YES	YES	YES	YES	YES
State Fixed Effects	NO	YES	YES	NO	YES	YES
Observations	864	864	849	864	864	849
R-squared	0.170	0.278	0.277	0.172	0.281	0.281

Notes: The dependent variable in all specifications is the number of policy experiments (waiver applications) in a given state and year. Specifications (1) and (4) include year fixed effects, all other specifications additionally include state fixed effects. The past governor vote margin measures the winner's distance in votes to the runner-up in the last gubernatorial election. The LD (lame duck) dummy is equal to one if the governor cannot be reelected. Standard errors clustered at the state level are shown in parentheses. *** p<0.01, ** p<0.05 and *p<0.1.

Initial Reputation and Experimentation. The second prediction we test is whether a governor with a high initial reputation is less likely to experiment (Table 3.5). As before, we include year fixed effects only in specifications (1) and (4), add state fixed effects for all other columns and additionally the age of the governor in specifications (3) and (6). The right-hand side of the Table 3.5 allows for a differential effect of lame ducks. Since reputational concerns should matter less for governors facing binding time limits, a high initial reputation should not be systematically correlated with the decision to experiment in the sample of lame ducks. Table 3.5 generally finds support for this prediction. The propensity to experiment declines with initial reputation in the full sample of all governors. When differentiating between governors who can and cannot be reelected, the negative effect survives only for reelectable governors which is in line with our prediction. The effect is again sizeable: a one standard deviation in initial reputation reduces experimentation during the waiver period by about 27%.[22] For lame ducks

[21] Using the coefficients from column (6), a one standard deviation of the interaction term quality*caseload (0.284) multiplied with the estimated coefficient (1.671) yields an effect of 0.475. Evaluation at the mean number of policy experiments during the TANF period (0.96) yields an effect of 49.43%.

[22] Using the coefficients in column (6), one standard deviation of the past vote margin (7.25) multiplied with the estimated coefficient (-0.0074) yields a decline of 5.4% (-0.0539). Evaluated at the mean number of policy

in contrast, initial reputation is not systematically related to the decision to experiment: The effect for lame ducks is represented by the sum of the vote margin coefficient and the interaction effect of vote margin and lame duck. It is close to zero or at least very small for all specifications (4), (5), or (6).

Overall, our findings suggest that reputational concerns influence the decision to experiment with policy rules before as well as after the US federal welfare reform in 1996. We next turn to the question whether reelection concerns also matter for the decision to revert a policy once adopted.

3.6.2 Governor Quality and Policy Reversals

In the model, low-quality governors face high reputational costs of reversing a policy experiment. Hence, if the gains from experimentation are high, low-quality governors prefer to stick to a policy experiment and hope that the experiment turns out to be successful in the second period. To test this prediction, we regress an indicator whether a governor reverses a policy previously adopted on the gain from experimentation, whether a governor is of low quality, and the interaction between the two variables.[23] The specifications are the same as before.

Table 3.6 shows that low quality governors are indeed significantly less likely to reverse a policy experiment when the net gains are high. Differences to lame duck governors seem to play no large role in the case of reversals. A one standard deviation increase in the gains from experimentation among low-quality governors reduces the likelihood of a policy reversal by about 36%.[24]

In sum, our findings thus show that reelection concerns matter both for policy experimentation and the decision to reverse a policy experiment later on. The evidence for both the AFDC and TANF period is consistent with our political agency model in which both voters and politicians are uncertain about the appropriate policy, and in which politicians take into account how their policy choices affect their reputation among voters.

---

experiments during the waiver period per year (0.2), we get a decline by 26.93%.

[23]Since the reversal indicator is binary, a Probit model might be more appropriate. We find similar results when using a Probit model instead.

[24]Using the coefficients in column (6), a one standard deviation of the interaction term $low\_quality * caseload$ (0.0115) multiplied with the estimated coefficient (-4.74) yields an effect of -0.055. Evaluated at the mean number of reversals (0.153) yields an effect of -35.63%.

Table 3.6: Electoral Incentives and Policy Reversals in Post-TANF Period (1996-2010)

	All Governors			Differential Effect for Lameducks		
	(1)	(2)	(3)	(4)	(5)	(6)
Low Quality * Caseload in State	-4.532**	-4.195*	-4.277*	-4.204*	-3.754	-4.740*
	(1.851)	(2.441)	(2.435)	(2.316)	(2.704)	(2.676)
Low Quality	0.0949**	0.0680	0.0852*	0.0899	0.0560	0.0945
	(0.0442)	(0.0452)	(0.0459)	(0.0607)	(0.0552)	(0.0603)
Caseload in State	1.417	-0.212	-0.611	0.768	-0.469	-0.108
	(1.151)	(3.023)	(2.954)	(1.651)	(3.370)	(3.470)
Governor Age			0.00268			0.00268
			(0.00199)			(0.00210)
LD * Low Quality * Caseload in State				-0.359	-0.863	1.309
				(3.624)	(3.806)	(4.351)
LD * Low Quality				0.0116	0.0221	-0.0309
				(0.0755)	(0.0734)	(0.104)
LD * Caseload in State				1.220	0.709	-1.183
				(1.794)	(1.981)	(2.819)
LD * Governor Age						0.000649
						(0.00137)
Year Fixed Effects	YES	YES	YES	YES	YES	YES
State Fixed Effects	NO	YES	YES	NO	YES	YES
Observations	666	666	666	621	621	621
R-squared	0.155	0.216	0.218	0.155	0.224	0.226

Notes: The dependent variable in all specifications is an indicator whether the state has reversed any policy rule in a given year or not. Specifications (1) and (4) include year fixed effects, all other specifications additionally include state fixed effects. The low quality governor indicator is equal to one if the number of years of political experience is below the median value. Caseload refers to the percentage of AFDC/TANF recipients in the population, lagged by one year. The LD (lame duck) dummy is equal to one if the governor cannot be reelected. Standard errors clustered at the state level are shown in parentheses. *** p<0.01, ** p<0.05 and *p<0.1.

3.6.3 Robustness Analysis

Ideology. So far, we have analyzed the decision to experiment and reverse welfare policy rules abstracting from conflicts of interest between voters or politicians. In the model, voters and politicians agree on which policy is best given the state of the world. However, politicians and voters might also differ in their assessment which policy is best and care about the policy rule implemented because of ideological considerations. During the US welfare reform, Republicans were typically more in favor of the decision to abolish welfare and encourage workfare. Hence, they might have been more willing to experiment with stricter rules than Democratic state governors. Somewhat surprisingly, the party of the governor is not related to the decision to experiment with a waiver (see the first column in Table 3.7).

Table 3.7: The Role of Ideology for Policy Experimentation

	(1)	(2)	(3)	(4)	(5)	(6)
Caseload in State * Rep. Governor	-0.345					
	(3.850)					
Caseload in State * Dem. Governor	0.558					
	(4.107)					
Quality * Caseload in State		0.519**	0.529**	0.510**	0.568**	0.583**
		(0.254)	(0.257)	(0.249)	(0.269)	(0.287)
Quality		-0.0183*	-0.0186*	-0.0181*	-0.0196*	-0.0200*
		(0.00935)	(0.00949)	(0.00920)	(0.0101)	(0.0107)
Caseload in State		-5.831	-5.801	-5.801	-7.673	-7.277
		(5.149)	(5.124)	(5.195)	(5.635)	(5.701)
Governor Age		0.000963	0.00104	0.000867	0.000304	0.000249
		(0.00155)	(0.00159)	(0.00156)	(0.00194)	(0.00198)
Governor Party Dummy		-0.0155				0.192
		(0.0380)				(0.135)
Governor Party Ideology			0.0609			0.381
			(0.0770)			(0.276)
Citizens Ideology				-0.000213		0.000820
				(0.00277)		(0.00381)
Democratic Presidential Vote Share					0.00746	0.00634
					(0.00605)	(0.00605)
Year Fixed Effects	YES	YES	YES	YES	YES	YES
State Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	880	846	846	846	758	758
R-squared	0.274	0.284	0.284	0.284	0.290	0.291

Notes: The dependent variable in all specifications is the number of policy experiments (waiver applications) in a given state and year. All specifications include year and state fixed effects. The quality of a governor is measured by years of political experience, see the data appendix for details. Caseload refers to the percentage of AFDC/TANF recipients in the population, lagged by one year. 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 last presidential election. Standard errors clustered at the state level are shown in parentheses. *** $p<0.01$, ** $p<0.05$ and * $p<0.1$.

To test the influence of ideology more systematically, the decision to experiment is regressed on governor quality, lagged caseload and its interaction effect (as in Table 3.3). In addition, we also control for several ideology measures: We include the party of the governor or a measure of governor ideology. We also proxy for the demand for welfare policy using a measure of citizen ideology (see Berry et al. (1998)) or the Democratic vote share in the last presidential election. Interestingly, none of these ideology measures matter for policy experimentation. Furthermore, the basic result that high-quality governors experiment more when the gains from experimentation are high remains statistically and economically significant.

Other Political Factors. A second feature of our analysis is that we focus on the governor as the sole political decision-maker. As we argue above, state governors played a prominent role during the waiver period and the passage of PRWORA. And yet, the legislature also has a say in welfare policy. Some evidence suggests that party composition and the closeness of the past race seem to have an effect on welfare spending under AFDC (e.g. Barrilleaux et al. (2002)). Here, we check whether the legislature also influences the decision whether a state experiments or not.

Table 3.8: Other Political Factors and Policy Experimentation

	(1)	(2)	(3)	(4)
Quality * Caseload in State	0.573**	0.503**	0.523**	0.565**
	(0.255)	(0.249)	(0.244)	(0.255)
Quality	-0.0201**	-0.0180*	-0.0191**	-0.0202**
	(0.00947)	(0.00915)	(0.00893)	(0.00937)
Caseload in State	-6.456	-5.536	-5.860	-6.259
	(5.260)	(5.193)	(5.151)	(5.347)
Governor Age	0.00103	0.000924	0.000868	0.00103
	(0.00177)	(0.00155)	(0.00157)	(0.00175)
Divided Government	0.0844**			0.0695*
	(0.0374)			(0.0407)
Dem. Seat Share Upper House		-0.270		-0.161
		(0.273)		(0.265)
Dem. Seat Share Lower House		-0.0356		-0.00224
		(0.316)		(0.321)
Polarization Upper House			-0.361	-0.220
			(0.350)	(0.345)
Polarization Lower House			-0.160	-0.0839
			(0.309)	(0.319)
Year Fixed Effects	YES	YES	YES	YES
State Fixed Effects	YES	YES	YES	YES
Observations	835	846	846	835
R-squared	0.288	0.285	0.286	0.289

Notes: The dependent variable in all specifications is the number of policy experiments (waiver applications) in a given state and year. All specifications include state and year fixed effects. The quality of a governor is measured by years of political experience, see the data appendix for details. Caseload refers to the percentage of AFDC/TANF recipients in the population, lagged by one year. 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%. Standard errors clustered at the state level are shown in parentheses. *** p<0.01, ** p<0.05 and *p<0.1.

Table 3.8 finds no evidence that the composition of the legislature (upper and lower house) or its polarization (often seen as an indicator for stiff political competition) have any effect on the decision to experiment. The only variable that is correlated with the decision to experiment is whether a state has a divided government. Somewhat surprisingly, we find that states with divided governments actually experiment more.[25] Most importantly, our conclusion that reputational concerns of the governor matter for the decision to experiment remains unchanged. Across all specifications in Table 3.8, the coefficient on the interaction effect remains statistically significant and very similar in size to the baseline specification in Table 3.3.

Table 3.9: Spillovers between States and Policy Experimentation

	(1)	(2)	(3)	(4)	(5)	(6)
Quality * Caseload in State	0.546**	0.561**	0.626**	0.528**	0.534**	0.577*
	(0.262)	(0.269)	(0.300)	(0.251)	(0.255)	(0.293)
Quality	-0.0201**	-0.0193*	-0.0222*	-0.0197**	-0.0186*	-0.0205*
	(0.00991)	(0.00999)	(0.0111)	(0.00950)	(0.00947)	(0.0106)
Caseload in State	-6.025	-6.717	-5.650	-6.965	-8.101	-6.926
	(5.443)	(5.344)	(5.939)	(5.227)	(4.950)	(5.444)
Governor Age	0.00135	0.000731	0.00230	0.00117	0.000325	0.00184
	(0.00179)	(0.00206)	(0.00203)	(0.00176)	(0.00198)	(0.00193)
Geographic Neighbors' Experiments (t-1)	-0.0978	-0.124	-0.124			
	(0.0840)	(0.0846)	(0.0880)			
Geographic Neighbors' Experiments (t-2)		-0.123	-0.117			
		(0.0877)	(0.0936)			
Geographic Neighbors' Experiments (t-3)			0.0554			
			(0.106)			
Pop. Size Neighbors' Experiments (t-1)				0.247	0.253	0.266
				(0.204)	(0.208)	(0.199)
Pop. Size Neighbors' Experiments (t-2)					0.0667	0.0779
					(0.158)	(0.161)
Pop. Size Neighbors' Experiments (t-3)						-0.0397
						(0.159)
Year Fixed Effects	YES	YES	YES	YES	YES	YES
State Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	802	758	713	802	758	713
R-squared	0.284	0.292	0.296	0.292	0.299	0.303

Notes: The dependent variable in all specifications is the number of policy experiments (waiver applications) in a given state and year. All specifications include year and state fixed effects. The quality of a governor is measured by years of political experience, see the data appendix for details. Caseload refers to the percentage of AFDC/TANF recipients in the population, lagged by one year. Geographic Neighbors' Experiments refers to the average value of the dependent variable for adjacent states (geographically neighboring states). Pop. Size Neighbors' Experiments refers to the average value of the dependent variable for states of similar population size (where all states are divided into ten bands of similar population size). Standard errors clustered at the state level are shown in parentheses. *** $p<0.01$, ** $p<0.05$ and * $p<0.1$.

Spillovers. Our model and analysis also abstracts from potential spillovers between 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

[25] For a detailed analysis of the effects of divided government on reform-making in US states, see Bernecker (2014).

neighboring states (see Brueckner (2000), Figlio et al. (1999), Saavedra (2000), or Wheaton (2000)). Spillovers could also arise if policy choices in neighboring states provide information for voters (as in a model of yardstick competition (Besley and Case (1995b)) or for politicians within a federation (as stressed by Oates (1999) or Strumpf (2002)).

To test for spillover effects during the US welfare reform, we create several measures of experimentation in neighboring states. Our first measure calculates how much geographic neighbors experiment on average in prior years. Alternatively, we define neighbors as states with similar population size and create their mean propensity to experiment in previous years. Table 3.9 shows that there is no systematic relationship between experimentation in a state and experimentation in neighboring states over the past three years. And as before, including controls for potential spillover effects between states does not change our baseline conclusion that reelection concerns matter.

Table 3.10: Socioeconomic Characteristics and Policy Experimentation

	(1)	(2)	(3)
Quality * Caseload in State	0.512**	0.519**	0.455*
	(0.245)	(0.247)	(0.251)
Quality	-0.0180**	-0.0179**	-0.0159*
	(0.00878)	(0.00886)	(0.00889)
Caseload in State	-8.558	-9.213*	-9.471*
	(5.167)	(5.151)	(5.334)
Governor Age	-0.000149	-0.000119	-0.000365
	(0.00159)	(0.00156)	(0.00147)
Unemployment Rate (/1000)	0.305	0.333	-2.295
	(12.52)	(12.46)	(11.66)
Per Capita Income (/1000)	-0.00311	-0.0119	-0.00326
	(0.0185)	(0.0195)	(0.0202)
Population (/1000)	0.106***	0.122***	0.167***
	(0.0318)	(0.0322)	(0.0390)
% Population Black		0.0480	0.0577
		(0.0604)	(0.0573)
% Population 65+		0.0313	0.0318
		(0.0425)	(0.0442)
% Immigrant Population			-0.0504***
			(0.0128)
Gini Coefficient (Family Income)			-0.252
			(1.056)
Year Fixed Effects	YES	YES	YES
State Fixed Effects	YES	YES	YES
Observations	846	846	846
R-squared	0.293	0.294	0.301

Notes: The dependent variable in all specifications is the number of policy experiments (waiver applications) in a given state and year. All specifications include state and year fixed effects. The quality of a governor is measured by years of political experience, see the data appendix for details. Caseload refers to the percentage of AFDC/TANF recipients in the population, lagged by one year. Robust standard errors clustered at the state level are shown in brackets. *** p<0.01, ** p<0.05 and *p<0.1.

Socioeconomic Characteristics. The empirical analysis so far further abstracts from the demand side. Time-varying economic or socio-economic conditions could shift the demand for welfare provisions and hence increase or reduce the desire to experiment. To test this prediction, we add state characteristics as proxies for the demand side to our baseline specification. The set of demographics controls we include are those commonly used in other studies of policy innovations (see, for example, Walker (1969), Berry and Berry (1992), or Boehmke and Skinner (2012)). Table 3.10 shows that states with a growing population experiment more, while states with a growing population of legal immigrants, in contrast, experiment less. We do not find any relationship between the decision to experiment and either the current unemployment rate, mean income per capita, income inequality (measured by the Gini coefficient), or the age and racial composition of the population. Our main result is again not affected by controlling for the demand side.

3.7 Conclusion

In this paper, we analyze how reputational concerns among governors influence the decision to experiment or reverse a policy experiment. Our empirical setting is the US welfare reform in 1996, the most dramatic shift in welfare policy since the New Deal. To motivate our empirical analysis, we outline a political agency model and derive several empirical predictions regarding how reelection concerns affect whether states experiment with or reverse welfare rules. To test our predictions, we build a novel data set of welfare policy experiments and reversals in the US states between 1978 and 2010. We complement our data with detailed information on the governor and legislature as well as additional socio-economic characteristics in each state.

Our predictions confirm that reelection concerns matter for the decision to experiment both during the waiver period before 1996, but also for the period after 1996 when states decide on their welfare policy themselves. Reelection concerns also matter for the decision to reverse an experiment during the TANF period. In line with our theoretical considerations, governors who face binding time limits ("lame ducks") behave systematically different from governors who face reelections and hence worry more about their political reputation. A battery of robustness tests suggests that policy innovations and reversals cannot be explained instead by ideological considerations, characteristics of the legislature, spillovers between states, or the demand for welfare services in a state.

Chapter 4

Do Politicians Shirk when Reelection Is Certain? Evidence from the German Parliament[1]

4.1 Introduction

Instead of working for the people who elected them into office, politicians might often shirk and invest their time in leisure activities or opportunities to earn outside incomes (on top of their payments as politicians). A recent example that has been heatedly discussed in the German media is former Minister of Finance Peer Steinbrück who had been nominated the Social Democratic Party's candidate for becoming Federal Chancellor of Germany after the September 2013 election. He had missed several important parliamentary sessions in order to give paid speeches elsewhere.[2] Another example is Silvana Koch-Mehrin, a former German Vice President of the European Parliament, whose low attendance rate at sessions of the European Parliament was repeatedly discussed in the media as well.[3]

Since voters cannot directly influence politicians' effort levels, they have to rely on voting in elections as the main device to keep politicians accountable.[4] The link between electoral

[1]For helpful comments and discussions, I thank Johannes Becker, Thomas Braendle, Christina Gathmann, Hans Peter Grüner, Friedrich Heinemann, Eckhard Janeba, Björn Kauder, Florian Misch, Steffen Osterloh, Andreas Peichl, Torsten Persson, Panu Poutvaara, and Johannes Rincke. I further thank seminar participants at the Annual Meeting of the German Economic Association, at the University of Mannheim, and at the Centre for European Economic Research. An earlier version of this chapter has been circulated in the University of Mannheim Economics Department Working Paper Series (Bernecker (2013)).

[2]For example, see Berliner Zeitung (2012).

[3]For example, see Hamburger Abendblatt (2011).

[4]For standard models of accountability, see Persson and Tabellini (2000) or Besley (2006).

competition and political performance therefore is a long-running theme in political economics.[5] One of the key questions is: Does harsher electoral pressure discipline politicians and reduce rent-seeking? This study provides micro-evidence on this topic by looking at election and new shirking data for politicians in Germany. The hypothesis is that stiffer electoral competition can discipline politicians for reelection motives: An MP (member of parliament) elected in her district with a very large vote margin can expect to be safely reelected in the next election even if her behavior in parliament is not exemplary. In contrast, an MP who was barely elected faces a higher risk that she might not be reelected next time. This might influence her rent-seeking behavior. In line with the examples given in the preceding paragraph, I operationalize rent-seeking by absence rates in parliament.

Regular debates about electoral reforms in a lot of countries demonstrate how relevant the analysis of the relation between electoral competition and shirking is. For instance, supporters of an electoral reform in Spain recently wanted to introduce an open-list system allowing voters "to pick and choose among individual candidates"[6] and in Britain's recent debate about the possible (and in the end rejected) introduction of the so-called "alternative vote" it was claimed that this would "force 'lazy' politicians 'to work harder' by reducing the number of safe seats"[7]. But in the end this is an open empirical question: How do politicians really react do different levels of electoral competition? I answer this question, also pointing towards implications for electoral reform. Existing evidence so far largely focuses on Italy which is arguably very different from Germany in terms of political system and culture. I also show in subsequent sections how different political mechanisms are relevant for Germany compared to Italy.

The basic setting of the paper is the following: Elected politicians derive utility from being in office, but not necessarily from taking part in parliamentary sessions. Therefore, a politician faces a trade-off between allocating time to political work (increasing her reelection probability) and allocating time to other non-political activities (such as earning outside incomes on top of the fixed MP payment or just relaxing, for example). Different MPs are confronted with different levels of electoral competition. This results in different levels of marginal utility of time allocated to political work: A politician running for reelection in a district with low electoral competition (where she can be almost certain to be reelected) has a pretty low marginal utility of investing additional time in political work and vice versa. This should be reflected in the time shares

[5] An example of an early contribution is Wittman (1989).
[6] See Economist (2011a).
[7] See Economist (2011b).

different MPs spend on political work and thus in MP absence rates.[8] Of course, there is also the theoretical possibility that an absent MP spends her time with voters in her district, potentially increasing her reelection probability this way. But empirically this seems not to be relevant, since the analysis reveals that absent MPs are indeed punished by voters in the next election. I also provide further evidence in favor of the rent-seeking interpretation of absences: MPs with high absence rates are at the same time the ones who do not answer their voters' questions asked at the leading political transparency website in Germany (www.abgeordnetenwatch.de).

The scope of this paper therefore is to estimate the effect of political competition on absence rates in parliament. The degree of political competition a politician faces is measured by her first vote margin in the past election (her own vote share minus the vote share of the runner-up in the same district) which is shown to be a proxy for the closeness of the coming election. A contribution of this paper is the construction of a unique measure for parliamentary absence rates in Germany using the recorded votes of the German parliament. The absence rate measures the share of mandatory parliamentary session days on which an MP misses to vote. In the analysis, I control for MPs who are excused (e.g. because of health issues), MPs who have a lot of other obligations (such as government ministers), and other factors. The baseline estimation is done by OLS. However, the analysis faces potential endogeneity issues. One example is that an MP could be highly motivated leading to a large vote margin and high session attendance at the same time. This type of omitted variable bias could lead the OLS coefficient to underestimate the effect of competition on absences. The analysis is therefore complemented by 2SLS estimations. The instrument used for the vote margin uses the special fact that in the German electoral system voters have two votes, one in the majoritarian tier of the system and another in the proportional tier of the system.

The key findings of the study are the following: There is a relevant and significant positive effect of vote margins on absence rates, i.e. MPs faced with low levels of electoral competition show higher absence rates in parliament. A closer look reveals that the effect is especially pronounced for opposition MPs while it is insignificant for MPs of governing parties. The explanation provided is that discipline enforcement by faction leaders is stronger within government parties since the government has to get its bills through parliament. For opposition MPs, an increase in the vote margin of 10 percentage points raises the absence rate by about 7 percentage points. This result is robust towards instrumenting the vote margin, measurement issues, and other factors. I extend the analysis of the effects of political competition to MPs

[8] See Becker et al. (2009) for a very similar argument made with respect to outside incomes of MPs.

elected via party lists and show that a safer list position implies a significantly higher absence rate.

The rest of the paper is organized as follows: The next section shortly reviews the related literature. Section 4.3 explains the institutional background and lines out the empirical strategy. Section 4.4 presents the data and the construction of the absence rates measure. Section 4.5 and Section 4.6 are devoted to results and robustness. Section 4.7 offers some extensions, Section 4.8 concludes.

4.2 Related Literature

There is a growing amount of empirical literature on the effects of political competition on economic outcomes in general. A first and very prominent example is Besley and Case (1995a). Investigating data about US governors from 1950 to 1986, they find that gubernatorial term limits significantly affect economic policy choices. An example of a more recent contribution is Besley et al. (2010) analyzing how a lack of political competition (measured by dominance of one party in state-wide elections) in US states can lead to policies that are harmful for economic growth. More relevant for this paper, however, is the strand of literature on the effects of political competition on political outcomes – on the aggregate and on the micro level.

Recent examples of studies that focus on aggregate political outcomes are Strömberg (2008), Dal Bó et al. (2009), or Svaleryd and Vlachos (2009). Strömberg (2008) analyzes the effect of political competition on campaign spending of US presidential candidates. Dal Bó et al. (2009) show that political dynasties are less likely to occur when political competition is more intense. Svaleryd and Vlachos (2009) illustrate that Swedish municipal councils are less likely to increase their wages and party funding when political competition between party blocs is larger. All these studies focus on political outcomes that are typically not solely decided upon by the individual politician.

In contrast, the literature on the micro level focuses on how individual politicians react to political competition by adjusting their rent-seeking behavior. A prominent recent example is Ferraz and Finan (2011) showing that Brazilian mayors who can be reelected are less corrupt. Another example is Snyder and Strömberg (2010) investigating how different levels of political competition induced by press coverage affect in versus against party line voting by members of the US Congress. Becker et al. (2009) analyze the effect of political competition on politicians' outside earnings, i.e. income earned on top of the MP salary. The latter study is close to mine in so far as it also looks at Germany and uses a similar methodology. They find that a decrease in the vote margin by 10 percentage points decreases the outside income by about 17,000 Euros per year.[9] Analyzing outside earnings is definitely a very natural and important first step, but likely not sufficient to fully analyze rent-seeking by MPs. An MP may, for example, also simply enjoy the sun in her garden instead of going to parliament. This is why I suggest complementing their analysis by using another measure. I also look at a considerably longer time horizon and include an analysis of list MPs.[10]

The rent-seeking measure employed in this study is absence rates from parliamentary sessions. This measure has been used most prominently to analyze MP behavior in the Italian and in the European Parliament.[11] For the European Parliament, Mocan and Altindag (2013) use a pay scheme harmonization reform in 2009 to investigate the effects of MPs' salaries on attendance behavior in parliament. They find that a salary increase reduces attendance. Using the same reform, Fisman et al. (2012) find no effect of salary changes on absenteeism. Both papers focus on MP payment and do not look at political competition on the MP level. For Italy, Gagliarducci et al. (2010) show that larger absence rates of MPs are associated with larger outside earnings. Gagliarducci et al. (2011) use a regression discontinuity design to illustrate that Italian MPs elected through the majoritarian tier of the system have lower absence rates than their colleagues from the proportional tier. One explanation could be that majoritarian tier MPs are faced with stiffer competition compared to proportional tier MPs. However, the study does not investigate competition effects within an electoral system. By now, there are also two studies looking at absence rates and other MP performance measures of German MPs: Arnold (2013) and Kauder and Potrafke (2014) both analyze the link between MP performance in parliament and MP outside earnings. While the former finds strong evidence for a relationship between outside earnings and absence rates, the latter do not. This may be explained by the fact that Arnold (2013) uses a selection on observables approach and a selected sample of recorded votes obtained from a German transparency platform, while Kauder and Potrafke (2014) uses a potential market income approach and the full set of recorded votes obtained from the German parliament website. In contrast to my paper, both of these papers do not focus on the role of political competition in determining absences.

[9] For a discussion of moonlighting in Germany, see Geys and Mause (2012). For a general survey on moonlighting, see Geys and Mause (2013).

[10] Another potential issue is that (unobservable) outside earnings opportunities may not be constant across MPs and may need some "reaction time" to unfold if they depend on an MP's reputation (and can therefore not adjust immediately to election results). Complementarity of absence rates to the measure of Becker et al. (2009) is supported by the fact that I find different results compared to theirs for MPs of governing parties.

[11] Lott (1987) and Besley and Larcinese (2011) are examples looking at the US and the UK respectively. None of these analyzes the role of political competition.

The two papers that are closest to my study are Galasso and Nannicini (2011) and Nannicini et al. (2013). Both again focus on the Italian case. Galasso and Nannicini (2011) present a model where party leaders allocate high quality candidates to more competitive districts and politicians elected in competitive districts are less absent in parliament. They check their hypotheses using data from 1994 to 2006 and find that the lower absence rates of MPs in competitive districts are more due to selection (high quality candidates are less absent) than to reelection motives (small vote margins encourage not to shirk too much). Galasso and Nannicini (2011) claim that high quality candidates (who are less absent anyway) are put in competitive districts by party leaders. However, in Germany who becomes district candidate of a party is not determined by party leaders, but by party conventions and thus more decentralized. I show how the German political setting is very different from the Italian one and that their result does not hold in general. In contrast to the finding of Galasso and Nannicini (2011) for Italy, reelection concerns seem to be the driving force of absences in Germany. Nannicini et al. (2013) show that in districts with higher social capital (measured by blood donations) MPs with high absence rates and criminal investigations are punished more by voters. Analyzing reelection chances of politicians facing criminal prosecution by itself already illustrates how the Italian case may be different from the German one (where it is hard to imagine such politicians to have any chance at all of being reelected).

To my best knowledge, this study is one of the first to create an absence rate measure for German MPs from parliamentary records. Abstracting from studies focusing on Italy, I am also the first to look at the effect of political competition on absence rates at the MP level. Italy seems to be very different in terms of political culture and system compared to Germany. My results confirm this conjecture and illustrate the need for studies looking at other countries also. This study is also the first to extend the analysis of the effects of political competition on MPs elected via party lists. Given the prevalence of proportional electoral rules in many countries, this seems to be a needed next step.

4.3 Institutional Background and Empirical Strategy

4.3.1 Institutional Background

Elections for the German parliament (Bundestag) usually take place every four years. The electoral system is mixed: It has a majoritarian (299 parliamentary seats) and a proportional tier (at least 299 additional parliamentary seats, usually slightly above this number) which results in a total of at least 598 seats in parliament. There are 299 electoral districts (constituencies). Each voter has two votes. The first one is used to vote for one of the candidates running in the district. In each district, the candidate gaining the most of these first votes is elected to parliament (first-past-the-post). This makes in total 299 so-called "direct" or "district" MPs. This is the majoritarian tier of the system. The second vote is to be cast for one of the parties running in the election. Here, the parties get seats according to their overall vote share (accumulated across districts). Who is sent to parliament by the parties is determined by "party lists" that are agreed upon at the state level in each party before the election. These additional MPs are the so-called "list MPs". This is the proportional tier of the system. The seats in the proportional tier are distributed in such a way that the parties' overall parliamentary seat shares reflect the distribution of second vote shares in the election. This is why usually the second vote is considered the more important one. There is a threshold such that only parties gaining at least a share of 5 percent of the second votes or winning at least 3 districts with their candidates get into parliament.[12]

In this study, I focus on the district MPs for the main part of the analysis and look at list MPs in an extension. Since in almost all cases a district winner runs again in the very same district in the next election, it is very reasonable to assume that a direct MP takes the closeness of the last election as a proxy for the closeness of the next one.[13] The higher an MP's first vote margin, the smaller is the level of electoral pressure she is faced with. I measure the closeness of a district race by the first vote margin: The first vote margin is the first vote share of the district MP minus the first vote share of the runner-up. I computed the vote margins from data obtained from the German electoral management board: For example, in constituency number 141 the two candidates who got most votes in the 2009 election were Axel Schäfer (43.3% of the first votes) from the Social Democrats and Norbert Lammert (31% of the first votes) from the Christian Democrats. This gives Axel Schäfer a first vote margin of 43.3 - 31 = 12.3 percentage points and made him a direct MP in the Bundestag. The vote margin is the first key variable in my analysis and the measure for electoral competition. Note that the Social Democrats and the Christian Democrats are the two large main parties in Germany. They win the vast majority of all districts and there has never been a chancellor from any other party. Most of my analysis

[12]To be exact, if a party fails to meet both these conditions, but wins 1 (or at maximum 2) districts directly, this one (or two) district winner(s) become MP(s) nevertheless, but the party does not get a whole faction in parliament according to its overall second vote share result.

[13]The within-candidate correlation of vote margins across elections is indeed high. For the large parties I focus on, it is about 0.65 and significant at any conventional level when looking at the elections of 2005 and 2009. See Table C.1 in the Appendix for a detailed analysis.

focuses on MPs of these two parties.

Since Norbert Lammert was placed safely on the Christian Democrats' party list, he nevertheless got into parliament through the proportional tier of the electoral system although having lost his district. This is an interesting feature of the German electoral system: Candidates can run for parliament in both tiers simultaneously. If they are elected in their constituency, they have to accept the majoritarian tier seat ("direct MP"), but if they do not win their constituency, they might still have a chance to become an MP if they are placed high enough on the party list ("list MP"). Usually, party conventions on the state level decide upon the party list for their state and almost all candidates who are put on the list also run for election as a candidate in one of the constituencies. It is mainly an MP's famousness and power within the party that determine her rank on the list. Yet, it is not quite clear how to determine the level of political competition that list MPs are faced with since this is mostly intra-party competition and hardly observable for voters and the researcher. This is why this study mostly looks at directly elected MPs. Nevertheless, I have a short extension on list MPs. I also exploit this peculiarity of the German electoral system – that candidates can run in both tiers simultaneously – to tackle potential endogeneity issues. I come back to this when describing the empirical strategy. It is also important to note that because of this feature of the system one has to control for party lists even in the analysis of district MPs: An MP with a first vote margin of only 1 percentage point might still not be under a lot of electoral pressure if she knows that she will be put on the first place of her party's list (and therefore very likely be elected to parliament through the proportional tier anyway).

This leaves us with the second key variable which is absence rates. I use the so-called "namentliche Abstimmungen" (mandatory votes recorded at the individual level) that can be called for by a number of at least 5% of all MPs or by a whole faction. All these votes are published in the official parliamentary records including those MPs who did not cast their vote with the remark "vote not casted". From this remark one can conclude that an MP did not take part in the vote, but was instead absent. Recorded votes cover all kinds of possible policies.[14] No committee sessions are scheduled parallel to recorded votes. Since recorded votes are typically called for in important cases and therefore all MPs are required to attend them, I expect absence rates at recorded votes to be rather small compared to overall MP absence rates in parliament. In that sense, my measure provides a conservative lower bound on parliamentary

[14]Recent examples include Bundeswehr assignments, minimum wages, the budget, the inheritance tax, the commuter tax relief, the Lissabon Treaty, data retention, corporate taxation, health care reform and an anti-discrimination law.

absences. While one may argue about whether absences in general reflect political shirking, it is safe to interpret absences as shirking in the case of recorded votes. Attendance is mandatory and the parliament administration is required by law to punish MPs for non-attendance via wage deductions. Missing a session day typically implies a wage deduction of 100 Euros.[15] I show in an extension that also voters indeed punish MPs with high absence rates in the next election.

4.3.2 Empirical Strategy

The hypothesis is that stiffer electoral competition leads to reduced political rent-seeking. The degree of competition an MP faces is measured by her vote margin in the past election. The assumption is that the past vote margin is a good proxy for the closeness of the upcoming electoral race. Table C.1 in the Appendix uses controlled regressions of the vote margin on the past vote margin to clearly support this assumption. Rent-seeking is operationalized via absences at mandatory sessions. Other performance measures such as the number of speeches given by an MP or special appointments in parliament are difficult to use since they are not under direct control of the MP. For example, the party faction as a whole or the faction leaders usually decide on whom to appoint to committee chairman or who should speak in the party's name on a certain topic. These decisions may also result from bargaining that is unobservable to the researcher. A problem with using bill proposals as measure is that they may in fact often be written by administrative staff. Data on attendance in committee sessions which could be another interesting measure is not available. This is why I rely on attendance in mandatory voting sessions. At least for the European parliament, there is also clear evidence that higher attendance rates result in increased legislative output at the MP level and thus indeed attendance captures MP effort.[16]

Estimation is done by OLS in the baseline:[17]

$$absences_i = \alpha + \beta * margin_i + controls_i * \gamma + \epsilon_i$$

where *absences* denotes the % of mandatory session days at which an MP was absent, *margin* the vote margin in the preceding election, and *controls* a number of other variables

[15] See §14(2) Abgeordnetengesetz. For a recent analysis of the payment of German MPs compared to non-politicians, see Peichl et al. (2013).

[16] See Fisman et al. (2012).

[17] Since the dependent variable is a percentage that is bound between 0 and 100, as a robustness check I also estimated generalized linear regressions with the share of absent days (between 0 and 1) as dependent variable assumed to be binomially distributed and using a logit link function (to take the limited range of the dependent variable into account). Results are very similar.

at the MP level. These include demographic characteristics such as gender and age, children, university, and PhD dummies, but also variables related to the political biography of an MP such as party membership, number of terms served in parliament, and minister and leading position dummies. A complete description of all variables including data sources is given in Table C.2 in the Appendix. If the hypothesis holds, one expects β to carry a positive sign: A large vote margin translates into a large expected vote margin in the next election, which means that electoral competition is expected to be weak resulting in high absence rates.

The reason for including dummies for ministers and leading positions (such as committee chairman) is that especially famous or influential MPs might gain large vote margins, but at the same time might obtain more special appointments and therefore have systematically less time to attend votes in parliament than other MPs. The number of terms served in parliament is included for similar reasons: An experienced MP might win her district with a relatively comfortable vote share, but might also change her behavior in parliament because of her experience or influence within parliament.

With the outlined approach, we face two potential endogeneity problems: First, imagine an MP to be very motivated. This might drive up her first vote margin, but also increase her attendance rate in parliament. Since one cannot control for unobserved motivation, OLS estimates could be biased (most likely downwards) due to omitted variable bias. Second, imagine a high vote margin indeed leading to a high absence rate which in turn – in the next election – may lead to a low vote margin via punishment by voters. In that case, we may be faced with a sort of equilibrium relationship between absences and vote margins in which MPs optimize their absences such that they get barely reelected. This could again result in a downward bias of the OLS estimates (due to simultaneity bias). Therefore, I also employ a 2SLS estimation where the potentially endogenous vote margin is instrumented. The first stage regression is:

$$margin_i = \delta + \zeta * secondvoteshare_i + controls_i * \eta + \mu_i$$

The share of second votes an MP's party gets in the MP's district is used to instrument for the vote margin of the MP (which is based on first votes in the district).[18] As explained before, the voters in a district give their first vote to a district candidate and their second vote to a party. The share of second votes of an MP's party in her district thus measures how many

[18]First and second votes are both collected at the district level. In that sense, the majoritarian and the proportional districts geographically coincide when it comes to the actual voting. However, party lists from which the list MPs are drawn are agreed upon by each party for a whole state.

people voted in favor of the MP's party in her district. This instrument has been used first by Becker et al. (2009). Since the first and the second votes are cast at the same time and in fact a lot of voters cast both votes according to party preferences, there is a high correlation between the two votes. If it also holds that voters' decision about how to cast the second vote is solely determined by party preferences, the instrument is not correlated with ϵ_i and is valid. This seems to be a reasonable assumption for the German electoral system. The second vote is informally even known as the "party vote" among the population. However, one can certainly come up with stories that would invalidate the instrument based on voters casting not only the first, but also the second vote based on preferences over the district candidates. I come back to this in the results section. In short, I construct yet another instrument based on earlier election results that cannot be driven by voter preferences over current district candidates.

Note that the 2SLS approach may itself underestimate the effect of interest: There is the general possibility that party conventions at the district level punish absent MPs by not nominating them again as their candidate. This would weaken the observed relationship between absences and vote margins. The 2SLS results therefore provide a conservative lower bound regarding the size of the effect.

4.4 Data

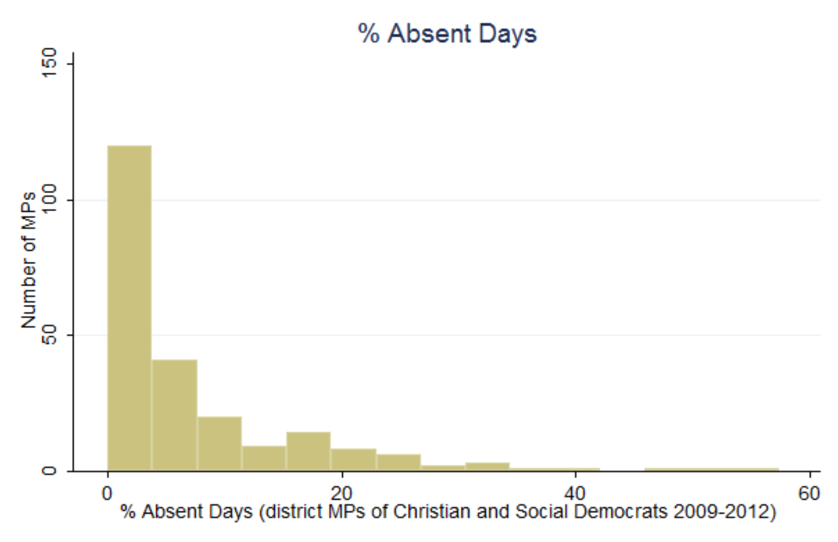
The period of analysis for the main part of the study is September 2009 (election of the parliament) until July 2012. In the parliament, there are 299 direct MPs, of which 289 have continuously been a member for the whole legislative period. Out of these, I focus on the MPs of the two large parties Christian Democrats and Social Democrats which leaves me with 228 observations for the baseline analysis. I exclude smaller parties since they usually only have very few MPs among the direct MPs which makes identification of party-specific effects (which turn out to be important) impossible.[19]

The dependent variable is a measure of MP absence rates in parliament. I constructed the absences for each MP using the data on recorded votes publicly available at the website of the German parliament. I include virtually all recorded votes that have been conducted from September 2009 up to July 2012 which gives me a total number of 141 votes taken on 53 different days.[20] The first recorded vote in the current parliament has been taken on the 3rd of December

[19]The main results remain unchanged when also including the largest of the smaller parties – the Christian Social Democrats – in the regressions. See Table C.3 in the Appendix.

[20]I exclude all 23 votes taken on 10/28/2010 because of data availability issues. This day should be excluded anyway since on no other occasion such a large amount of very similar votes has been called for. According to

Figure 4.1: Distribution of % Absent Days in German Parliament



2009, the last (up to the mid of July 2012) on the 29th of June 2012. Since there are up to 9 recorded votes per day, but sometimes only 1 or 2, I compute absence rates on a daily basis not to get my results biased or driven by single days: An MP is classified as absent on a certain day if she missed one or more of the recorded votes on that day. I divide the resulting number of missed days by the total number of days with recorded votes (53) to get the share of missed days for the whole legislative term up to July 2012.[21] Of course, there are other ways to measure absences. Later checks show that my results are robust to using different absence measures. As outlined before, I expect this measure to be a lower bound on parliamentary absences. Figure 4.1 shows the distribution of absences across MPs. As expected, absence rates are not too high for many MPs. The majority actually has an absence rate lower than 5%. On the other hand, there are MPs who missed more than 50% of all mandatory sessions and less than 17% of all MPs have never missed a session. Thus, there is quite some variation in MP absence rates. The average is at about 8%, the standard deviation is about 9%.

The main explanatory variable is an MP's vote margin as a measure of electoral competition. The average vote margin achieved is about 11 percentage points, but the standard deviation is more than 8 percentage points. Thus, there is also relevant variation in the level of political competition across MPs. There are MPs who have won the district with more than 43 percentage points ahead of the runner-up, but also others who have barely captured victory.

The analysis also includes a range of demographic and political controls. About 21% of the

the Bundestag administration, this is also why the data are not fully available for this day.

[21] There is no loss of information resulting from the aggregation over the years since checks reveal that there is no such thing as a political business cycle in absences over the legislative period.

MPs are female. The average age in 2012 is 53 years. Almost two thirds of the MPs have at least one child. More than 80% have attended university, more than 20% have a PhD. For the average MP the current legislative period is his or her third one served already. About one quarter of the MPs has taken a leading position in the current parliament (such as committee chairman or faction leader). For summary statistics and a full list of controls, see Table C.2 in the Appendix.

4.5 Results

OLS. Table 4.1 presents the OLS baseline results. The dependent variable is the percentage of absent days. Specification (1) includes the vote margin as main explanatory variable and a Social Democrats dummy (Christian Democrats is the base category). The vote margin has a highly significant effect: A 10 percentage points higher vote margin implies a 2 percentage points higher absence rate. Being a Social Democrat increases your absence rate by about 8 percentage points compared to a Christian Democrat. Since both effects are significant, specification (2) checks an interaction term which allows the vote margin effect to be different for Social versus Christian Democrats. It turns out that while Christian Democrats' absences seem not to react to electoral competition, Social Democrats are strongly responsive to it: For a Social Democrat, a 10 percentage points larger vote margin implies an absence rate which is about 7 percentage points larger. In this specification, the main effect of vote margin and the Social Democrats dummy are no longer significant. The effect is stable in terms of size and significance when adding several demographic and political controls on the MP level in specifications (3) and (4). Social Democrats who are confident that they will be reelected attend fewer sessions. I argue that the non-effect for the Christian Democrats is related to the fact that this party was in government at that time. Also in Italy MPs of governing parties are less absent.[22] I get back to this in Section 4.7.

In the most general specification all demographic characteristics are insignificant. Among the political controls, the minister dummy and the number of terms served are significant: Being a government minister increases your absence rate by about 11 percentage points, serving as an MP for one more legislative period increases it by about 1 percentage point. The former finding is easily explained by the fact that members of the executive have many obvious duties to perform outside the parliament which is why they cannot attend sessions. The latter finding fits the vote margin finding quite nicely: MPs who win their districts repeatedly are naturally

[22] See, for example, Galasso and Nannicini (2011) or Gagliarducci et al. (2010).

Table 4.1: Effect of Vote Margin on Absent Days (OLS)

Dependent variable: % Absent days	(1)	(2)	(3)	(4)
Vote margin * Social Democrat		0.664**	0.664**	0.733**
		(0.267)	(0.271)	(0.288)
Vote margin	0.192***	0.0994	0.104	-0.0172
	(0.0714)	(0.0714)	(0.0675)	(0.0562)
Social Democrat	8.115***	3.010	3.070	2.183
	(1.602)	(1.907)	(2.016)	(2.119)
Female			2.474	1.871
			(2.003)	(1.734)
Age			0.158**	-0.00214
			(0.0654)	(0.0686)
Partner			2.647	2.679
			(2.041)	(1.921)
Children			-2.382	-1.644
			(1.879)	(1.681)
University			2.768**	1.290
			(1.239)	(1.245)
PhD			3.301**	1.863
			(1.660)	(1.427)
Number of terms				0.984***
				(0.333)
Minister				11.16***
				(4.123)
Leading position				2.476
				(1.513)
Observations	228	228	228	228
R-squared	0.138	0.177	0.252	0.352

Notes: Sample includes district MPs of Social Democratic and Christian Democratic Party of the parliament elected in 2009 (excluding early leavers and late entrants). Dependent variable is the percentage of days when the MP missed at least one recorded vote (out of all days with a recorded vote). All regressions include a constant. Robust standard errors are reported in parentheses. *, **, *** denote significance at the 10, 5, 1 % level.

very confident that they will also succeed the next time. Therefore, over time they attend fewer and fewer sessions.

Table 4.2: Effect of Vote Margin on Absent Days (2SLS)

Dependent variable: % Absent days				
	(1)	(2)	(3)	(4)
Instruments taken from election year:	2009	1990	1990	1990
MP sample:	all	all	not run 1990	not run 1990 no redistricting
Vote margin * SD	0.809***	0.867**	0.737***	0.938***
	(0.307)	(0.367)	(0.260)	(0.223)
Vote margin	-0.0473	-0.0410	-0.0146	-0.0498
	(0.0655)	(0.0706)	(0.0639)	(0.0628)
Social Democrat (SD)	1.465	0.842	1.475	-2.435
	(2.401)	(2.830)	(2.225)	(2.620)
Demographic Controls	YES	YES	YES	YES
Political Controls	YES	YES	YES	YES
Observations	228	218	193	124
R-squared	0.351	0.364	0.286	0.279
1st stage regression for Vote margin				
F statistic	35.27	30.30	23.26	26.11
coefficient of IV Second vote share	1.467***	1.046***	1.0178***	1.152***
	(0.0949)	(0.0741)	(0.0811)	(0.0863)
*1st stage regression for Vote margin * SD*				
F statistic	73.78	78.85	74.24	56.34
coefficient of IV Second vote share * SD	0.981***	0.669***	0.636***	0.702***
	(0.0669)	(0.0451)	(0.0451)	(0.0542)

Notes: Sample includes district MPs of Social Democratic and Christian Democratic Party of the parliament elected in 2009 (excluding early leavers and late entrants) in column (1). In column (2), some MPs are dropped due to data availability issues in 1990. In column (3) sample size drops further, since all MPs who already run in the 1990 election are excluded. In column (4) sample size drops yet further, since all MPs whose districts have undergone relevant redistricting between 1990 and 2009 are dropped. Dependent variable is the percentage of days when the MP missed at least one recorded vote (out of all days with a recorded vote). All columns report 2SLS regressions. The instruments for vote margin and vote margin * Social Democrat are second vote share and second vote share * Social Democrat, respectively. In column (1), the second vote share is taken from the 2009 election, in columns (2) to (4) it is taken from the 1990 election. Key results from the two first stage regressions are always shortly reported in the lower part of the table. The first stage regressions are fully reported in the Appendix. Controls are the same as in Table 1 (demographic: female, age, partner, children, university, phd; political: number of terms, minister, leading position). All regressions include a constant. Robust standard errors are reported in parentheses. *, **, *** denote significance at the 10, 5, 1 % level.

2SLS. As outlined before, the OLS estimates in Table 4.1 may be plagued by endogeneity issues. This is why Table 4.2 presents results from 2SLS regressions where the second vote share achieved by the MP's party in her district is used to instrument for her vote margin. To be exact, the vote margin is instrumented by the second vote share and the interaction of vote margin and Social Democrat is instrumented by the interaction of second vote share and Social Democrat. The four columns all present 2SLS versions of the baseline OLS regression presented in column (4) of Table 4.1, but differ from one another by using different types of the instrument and different samples. Column (1) uses the second vote share from the 2009 election to instrument for the 2009 vote margin. The instrument is indeed strong, as can be seen from the key results of the two first stage regressions reported in the lower end of Table 4.2. The full first stage regressions can be found in Appendix Table C.4. In the 2SLS regression, the coefficient of main interest (Vote margin * SD) is 0.809 and highly significant. This number is slightly larger than the coefficient of 0.733 obtained when using OLS (column (4) of Table 4.1). This is in line with the expected downwards bias in OLS esimates described in Section 4.3.2).

As explained before, one potential caveat regarding the second vote share instrument may be that people not only cast the first vote, but also the second vote according to candidate preferences invalidating the instrument. This is why column (2) uses election results from the past to instrument for the current (2009) vote margin. The election results from 1990 are the first ones after German reunification and thus the earliest that can be used.[23] The first vote margin of a district MP is now instrumented by the second vote share the MP's party obtained in this district in the 1990 elections. This rules out that candidate-specific effects potentially affect the instrument. Column (2) shows that the 1990 instrument is indeed strong and that the main result is basically unchanged. Columns (3) and (4) address further potential caveats by repeating the analysis from column (2), but restricting the sample analyzed. First, some of the MPs elected in 2009 may have run in their districts in 1990 already. In this case candidate-specific effects may actually influence even the 1990 instrument. This is why column (3) excludes all MPs who did already run in 1990 from the analysis. Second, due to redistricting between 1990 and 2009 some of the assignments of 1990 district election results to 2009 district election results may be debatable. This is why column (4) cuts the sample further down by focusing on cases without any relevant redistricting. As can be seen in Table 4.2, both checks do not relevantly change the result from before in terms of size or significance of the effect.

Thus, also when taking into account endogeneity via 2SLS regressions, the results from before remain unchanged: The more certain a Social Democrats MP is of being reelected the fewer sessions she attends. For simplification and since the OLS estimates if anything slightly underestimate the effect of interest, I focus on OLS regressions in the next sections.

4.6 Robustness

Absences Measurement. In this section, I explore the robustness of the results presented in the previous section. The first three specifications of Table 4.3 explore the measurement of the dependent variable. In the baseline from the previous section (replicated in specification (1)),

[23] I follow Becker et al. (2009) here.

Table 4.3: Effect of Vote Margin on Absent Days (Robustness)

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Absent days	Absent votes	Absent w/o excuse	Absent days	Absent days
Vote margin * Social Democrat	0.733**	0.646**	0.285*	0.665**	0.758***
	(0.288)	(0.258)	(0.167)	(0.281)	(0.277)
Vote margin	-0.0172	-0.0259	-0.0342	0.0453	-0.0203
	(0.0562)	(0.0548)	(0.0379)	(0.0874)	(0.0554)
Social Democrat	2.183	1.192	1.770	2.700	2.822
	(2.119)	(1.962)	(1.305)	(2.145)	(2.053)
Safe				-0.847	
				(1.782)	
Vote margin * safe				-0.0708	
				(0.121)	
% Platform Answers					-0.0463***
					(0.0141)
Demographic Controls	YES	YES	YES	YES	YES
Political Controls	YES	YES	YES	YES	YES
Observations	228	228	228	228	228
R-squared	0.352	0.325	0.334	0.360	0.385

Notes: Sample includes district MPs of Social Democratic and Christian Democratic Party of the parliament elected in 2009 (excluding early leavers and late entrants). Dependent variable is the percentage of days when the MP missed at least one recorded vote (out of all days with a recorded vote) in specifications (1), (4), and (5). In specification (2), the dependent variable is the percentage of recorded votes the MP missed. In specification (3), the dependent variable is alike the one in specification (1), but does not count days as absent when the MP excused herself. Safe denotes that the MP would have been elected to parliament via the party list if she had lost her district. % Platform answers refers to the share of answered questions at the German transparency platform abgeordnetenwatch.de where citizens can ask MPs all kinds of questions. Controls are the same as in Table 1 (demographic: female, age, partner, children, university, phd; political: number of terms, minister, leading position). All regressions include a constant. Robust standard errors are reported in parentheses. *, **, *** denote significance at the 10, 5, 1 % level.

I use the absence rate on a daily basis not to get my results biased. Let me illustrate what is exactly meant by that by providing an example: MP X misses 9 recorded votes (that possibly all take place within the same three hours) on day A, but attends the (single) vote on day B, the (single) vote on day C, and the (single) vote on day D. MP Y on the other hand attends the 9 votes on day A, but misses the votes on days B, C, and D. If you count absolute absences (meaning votes instead of days), MP X has an absence rate of 9 out of 12 and MP Y has an absence rate of 3 out of 12. This way of measuring absences would clearly not capture what we want to get at: MP X was away for three hours, MP Y for three days, and still MP X has an absence rate that is three times larger than the one of MP Y. With daily measurement I try to (admittedly not perfectly) take that into account: In that case, the resulting absence rates would be 1 out of 4 for MP X and 3 out of 4 for MP Y. This is probably a better measure than using the absolute absences. Nevertheless, one could argue that this is an equally bad measure that just falsifies in the other direction. To show that the result is not dependent on the daily measurement of absences, I use the absolute absences (i.e. absent votes) as the dependent variable in specification (2) in Table 4.3. The results are very similar compared to before.

Another potential problem regarding the absence measure is how to deal with excuses. Along with the debate records, the parliament administration publishes a list of excused MPs for each session day. One could easily argue that excused absences should not be counted as absences in the sense I do: An MP who had a heart attack, lies in hospital and therefore misses a recorded vote can hardly be considered shirking. The average absence rate is down from 7.8% to about 4.5% if you take only unexcused absences into account. But there are severe problems caused by a simple exclusion of the excused absences: MPs are in fact allowed to excuse themselves for any (!) reason. And even more striking: They do not even have to announce the reason to the parliament administration when they want their names to be put on the list of excused MPs. It is thus easily possible to have your name put on the list without reporting any reason and enjoy some leisure time or give a paid speech somewhere else. For example, Peer Steinbrück who was mentioned as example in the introduction, excused himself from sessions when instead he talked elsewhere. And indeed, the published list of excused MPs does usually not give any reasons for the MPs' absences.[24] Potentially for that reason, the parliament administration reduces the MP wage even in the case of excused absences.[25] Overall, it can therefore be considered pretty arbitrary which names are on the list of excused MPs and I do not expect to cleanly identify the effect of the vote margin on absences when using the unexcused absences only. Specification (3) of Table 4.3 uses the percentage of unexcused absent days as dependent variable. Maybe surprisingly, even in this case there is still a significant vote margin effect for Social Democrats. It is of about half the size compared to before which makes sense since the average absence rate is of about half the size compared to before now as well. Thus, even when using this noisy measure you can find the disciplining effect of narrow vote margins on absence rates.

Safe Seats. Another concern with the analysis presented in the previous section is the following: If candidates in the German system can not only get into parliament by winning a district (majoritarian tier), but also by being put safely on a party list (proportional tier), an MP who won her district in a close race and does not have a safe place on her party's list might be systematically different from an MP who also won her district, but would have been elected to parliament anyway because of a safe place on the list: The degree of electoral pressure the two MPs are faced with is presumably way higher for the first one, since the latter is aware of the fact that she would have become an MP also in the case of a defeat in her district (which

[24] On very rare occasions there is a comment like "attended parliamentary convention of the Western European Union" or something similar.

[25] If you are on the list of MPs having excused themselves, the wage deduction drops from 100 Euros for a missed session day to 50 Euros. If you prove that you are in hospital, it drops further to 20 Euros. See §14(2) Abgeordnetengesetz.

is not true for the former). If we assume that candidates who have been placed safely on a party's list in 2009 will also be placed safely in the next election, even a pretty narrow margin of victory in a district might not be a strong incentive to work a lot and to attend all the votes in parliament. This may lead to a biased estimation (most likely underestimation) of the effect I am investigating.

Therefore, I include a dummy variable called "safe" as an additional regressor (also interacted with vote margin) in specification (4) of Table 4.3. The dummy is assigned a value of one if – all other results of the 2009 election held constant – the MP would also be a member of parliament now in the case of a defeat in her district and zero otherwise. This is the right way to define this indicator if we assume that the results of the 2009 election are taken as proxies for the next election. In fact, neither the dummy nor the interaction with vote margin are significant and also all the other coefficients barely change compared to baseline specification (1). Thus, bias via party lists seems not to be of any concern for the analysis. One explanation may be that there are special benefits from being a district MP over being a list MP such as larger independence from your party. Thus, even if a candidate expects to be elected to parliament via the party list she may still find it worthwhile to fight for winning her district. I come back to list MPs in particular in the next section.

Motivation. Absence rates are a very particular way of measuring whether an MP is working or not. In fact, one may argue, absent MPs may spend their time in their district and work on issues that are in the interest of their voters. I show evidence in the next section that this is very likely not the case since voters actually indeed punish MPs with high absence rates in the next election. As another check, I include a measure of "motivation" of the MP in specification (5) of Table 4.3 as another control. The measure chosen is the percentage of answered questions at the online transparency platform "abgeordnetenwatch.de" where citizens can ask their MPs all kinds of questions. A high answering percentage presumably indicates that an MP is highly active in trying to understand the worries of her voters and willing to work on the issues. In fact, this measure of MP motivation has a significant negative effect on absences in parliament: A 10 percentage points higher answering share is related to an absence rate that is about 5 percentage points lower. This shows that MPs with high absence rates are at the same time the ones who do not show a high motivation in terms of answering the questions of citizens.

Selection. A final concern may be that MPs' absence rates are not driven by reelection incentives but by selection instead: In a model tailored to the Italian case, Galasso and Nannicini (2011) show that parties aim at winning contestable districts by putting the "good" candidates (who also have fewer absences) into exactly these districts. They find evidence in this flavor for the Italian parliament. In that case, MPs in contestable districts (i.e. with small vote margins) do not show few absences because of the high electoral competition, but instead are MPs who are not often absent anyway and just happen to have been put into a contestable district by the party leader. – This mechanism is, however, not applicable to the German case. In Germany, the recruitment of district candidates follows very much local patterns. It is local party conventions nominating the candidates. An example illustrating the power of these local conventions that was discussed in the German media at that time comes from the 2009 elections: In a Hamburg district, the local members of the Social Democrats decided to nominate a newcomer as their candidate although both the party leader of the Social Democrats and the foreign minister clearly supported the current MP and (unsuccessfully) tried to intervene in his favor.[26] There is no way for German party leaders to allocate candidates all over Germany in a centralized fashion according to some "optimized master plan". Another fact is that in Germany MPs do typically not switch districts at all, but stay within one electoral area for their whole political careers.[27]

4.7 Extensions

Punishment for Absences. This section explores data base extensions to absence rates for list MPs and MPs of the parliament elected in 2005. When including absence rates for the previous parliament, one obvious issue to check is whether MPs who have been a member of both parliaments – the one elected in 2005 and the one elected in 2009 – are indeed punished by voters for having been absent at many sessions. This is what Table 4.4 investigates. The dependent variable is the vote margin in the 2009 election. The explanatory variable of interest is the absence rate in the previous legislative term (i.e. absences in the parliament elected in 2005). When standard controls are included, absences indeed have a significant negative effect on the vote margin in the next election, i.e. voters punish absent MPs by giving them fewer votes. A 10 percentage points larger absence rate in the legislative period 2005-2009 reduces an MP's vote margin in the 2009 election by between 1 and 2 percentage points. This again supports the suspicion that MPs do not necessarily spend the time being absent from sessions with voters or working for them.

Parliament Elected in 2005. Table 4.5 repeats the whole baseline analysis of the effect of

[26]See Spiegel (2008).

[27]For example, taking into account redistricting, not a single of the directly elected MPs has switched districts between 2005 and 2009.

vote margins on absence rates for the parliament elected in 2005.[28] For the parliament elected in 2009, the result was a significant and relevant effect for Social Democrats MPs. There is no reason to expect Social Democrats per se to be in general systematically very different in terms of sessions attendance from Christian Democrats. Therefore, I argued before that the non-finding for Christian Democrats can instead be explained by the fact that they are part of the government coalition in the parliament elected in 2009. If this is really the case, this should be reflected as well in the findings for the parliament elected in 2005: From 2005 to 2009 there was a so called grand coalition – consisting of both large parties Christian AND Social Democrats – governing. If being in government indeed turns the competition–absence channel off, we should therefore not find any effect of competition on absences for the parliament elected in 2005. This is indeed what Table 4.5 shows: Repeating the exact same type of analysis conducted before for the parliament elected in 2009 does not reveal any significant effect of vote margins on absence rates. I argued before that being part of the government coalition may have a disciplining effect on MPs, e.g. since there is an obligation to get the government's bills through parliament. MPs of governing parties also seem to have lower absence rates in general: The average absence rate of the Social Democrats increases from 11% to 13% when comparing the two legislative periods (of which only in the first the Social Democrats were part of the government). Another example of this is the Liberal Democrats who were in government in the 2009-2013 legislative period, but have not been in the previous one. For them, the average absence rate has gone down from 15% to 7%.[29]

Why is the competition–absence channel turned off for government MPs? – The answer may be that in a parliamentary democracy such as Germany the head of the executive is elected by the majority of MPs in parliament (the government party or a coalition of government parties). In Germany, this majority bringing the chancellor to power even has its own name which is "chancellor majority". The chancellor can rely on this majority to get her bills through parliament. In case a bill proposed by the government fails in parliament, this is therefore taken as an indication that the chancellor has lost the support of her own MPs. The media closely follows this. Thus, the government has every incentive to make its MPs being present at voting sessions. Party whip may play a role. But also to the individual MPs of governing parties it is clear that not attending voting sessions may substantially harm the own party. They may

[28]I again include virtually all recorded votes that have been conducted: These are 151 votes taken on 70 different days. The first one has been taken on November, 8th in 2005, the last one on September, 8th in 2009. I have to exclude 05/28/2009 and 05/29/2009 because of data availability issues again.

[29]The MPs of the Liberal Democrats are not included in the previous analysis since it only includes district MPs and the Liberals usually only have list MPs.

Table 4.4: Effect of Past Absent Days on Vote Margin

Dependent variable: Vote margin in the 2009 election	(1)	(2)	(3)
% Absent days 2005-2009	-0.109	-0.202***	-0.129**
	(0.0755)	(0.0681)	(0.0510)
Female		-5.393***	-3.566***
		(1.933)	(1.283)
Age		-0.183*	-0.0864
		(0.108)	(0.0753)
Children		1.694	-0.715
		(1.912)	(1.306)
University		-1.420	-0.930
		(2.252)	(1.866)
PhD		-2.719	-2.241
		(2.863)	(2.238)
Number of terms		1.623***	0.821*
		(0.506)	(0.424)
Minister		15.25***	10.20***
		(4.569)	(3.291)
Leading position		3.740	3.741**
		(2.329)	(1.730)
Social Democrat			-10.59***
			(1.284)
Christian Social Democrat			9.158***
			(1.740)
Observations	169	169	169
R-squared	0.011	0.217	0.594

Notes: Sample includes district MPs of Social Democratic, Christian Democratic, and Christian Social Party who have been members of both the parliament elected in 2005 and the parliament elected in 2009 (excluding early leavers and late entrants). Dependent variable is the vote margin in the 2009 election. All regressions include a constant. Robust standard errors are reported in parentheses. *, **, *** denote significance at the 10, 5, 1 % level.

therefore attend sessions irrespective of the individual level of competition they are faced with in their districts. MPs of opposition parties, on the other hand, do not have to show support of a government and have the freedom to respond to their own district competition levels when it comes to session attendance in parliament. This explanation is also consistent with lower absence rates in general for MPs of governing parties which have also been found for other countries.[30]

Since being in government or not thus matters a lot for attendance behavior in parliament, analyzing the 2005 and the 2009 parliament jointly would not yield much new insights. While in principle looking at two parliaments allows within-MP identification of the competition–absence channel via the time dimension, one would here need time variation in the vote margins of opposition MPs. But since the Social Democrats have been a governing party in the 2005 parliament and the Christian Democrats have been a governing party in both the 2005 and the

[30] See, for example, Galasso and Nannicini (2011) or Gagliarducci et al. (2010) for the Italian case.

Table 4.5: Effect of Vote Margin on Absent Days (Parliament Elected in 2005)

Dependent variable: % Absent days	(1)	(2)	(3)	(4)
Vote margin * Social Democrat		0.105	0.0850	0.209
		(0.146)	(0.142)	(0.130)
Vote margin	0.0626	-0.00633	0.0318	-0.0872
	(0.0722)	(0.113)	(0.113)	(0.105)
Social Democrat	1.209	-0.00335	-0.262	-0.883
	(1.181)	(2.020)	(1.967)	(1.838)
Demographic Controls	YES	YES	YES	YES
Political Controls	YES	YES	YES	YES
Observations	239	239	239	239
R-squared	0.007	0.009	0.071	0.301

Notes: Sample includes district MPs of Social Democratic and Christian Democratic Party of the parliament elected in 2005 (excluding early leavers and late entrants). Dependent variable is the percentage of days when the MP missed at least one recorded vote (out of all days with a recorded vote). Controls are the same as in Table 1 (demographic: female, age, partner, children, university, phd; political: number of terms, minister, leading position). All regressions include a constant. Robust standard errors are reported in parentheses. *, **, *** denote significance at the 10, 5, 1 % level.

2009 parliaments, this is infeasible. Another potential problem with such an analysis would be that there is quite some fluctuation in parliament membership across elections which would lower the number of observations considerably.[31] Thus, to allow for within-MP identification of the competition–absence channel in the presence of different government coalitions across electoral cycles one would need data for many more than just two parliaments. Absence data for the period before 2005 is however not available.[32]

Party List MPs. Table 4.6 finally presents a very first extension of the analysis to list MPs of the parliament elected in 2009. Besley et al. (2013) seems to contain the first model of party list construction. But they focus on gender composition of party lists and the Swedish case. In Sweden, party leaders are the ones composing the lists. In contrast, in Germany typically there is a separate vote at party conventions for each position on the list (with often several candidates running for the same position). To my knowledge, this extension is the first work to analyze the effect of list position on MP behavior. As explained before, list MPs do not get into parliament via being directly elected in a district, but by being put high enough on a party list (proportional tier of the system). This implies that these MPs are confronted with a different type of electoral

[31]Out of the 228 MPs analyzed for the 2009 parliament only 128 have already been members of the 2005 parliament. When pooling the data for the 2005 and 2009 parliaments nevertheless and including individual MP fixed effects, one still finds a significant positive effect of vote margins on absence rates for opposition MPs (not reported). This is consistent with the cross-sectional evidence presented in the main analysis of this paper.

[32]From 2005 onwards, the Bundestag administration has published lists of recorded votes including MP voting behavior on its website. For the period before, one would need to dig deep into the overall records of every plenary session of the parliament in order to identify recorded votes. Tremendous effort would be needed to achieve this even for one parliament already.

competition – which is within party competition for a promising position on the party list. While the first list position very often guarantees an MP being elected to parliament, positions further down on the list make it less and less likely that the party can grab enough votes for you to have a real chance of getting into parliament. Although there are differences across states and parties in which list position still ensures election to parliament, an MP's position number should in general be a very good proxy of the likelihood of election. Therefore, this extension uses an MP's list position in the preceding election to measure the level of political competition she is faced with, i.e. a list position with a low number (i.e. on top of the list) usually implies safe reelection while a position with a high number (i.e. further down on the list) implies that an MP has to worry about reelection.

Table 4.6: Effect of Party List Position on Absent Days

	(1)	(2)	(3)	(4)	(5)
Dependent variable:		Absent days			Absent w/o excuse
Party list position	-0.316***	-0.324***	-0.226**	-0.228**	-0.195***
	(0.0997)	(0.110)	(0.113)	(0.113)	(0.0658)
Social Democrat	4.422**	4.963**	5.446**	5.316**	1.171
	(2.007)	(2.185)	(2.100)	(2.178)	(1.694)
Liberal Democrat	0.395	0.577	1.389	1.274	-0.618
	(1.817)	(1.840)	(1.746)	(1.810)	(1.493)
Green party member	1.723	1.599	2.936	2.827	-0.0328
	(1.905)	(1.953)	(1.879)	(1.931)	(1.520)
Leftist party member	8.101***	8.060***	9.872***	9.771***	2.450
	(2.122)	(2.258)	(2.275)	(2.301)	(1.645)
Demographic Controls	NO	YES	YES	YES	YES
Political Controls	NO	NO	YES	YES	YES
% Platform answers				0.632	
				(1.771)	
Observations	309	258	258	258	258
R-squared	0.150	0.166	0.188	0.189	0.132

Notes: Sample includes all party list MPs of the parliament elected in 2009 (excluding early leavers and late entrants). Dependent variable is the percentage of days when the MP missed at least one recorded vote (out of all days with a recorded vote) in specifications (1) to (4). In specification (5), the dependent variable is alike the one in specification (1) to (4), but does not count days as absent when the MP excused herself. Party list position is the MP's number in the party list ranking of her party in her state. % Platform answers refers to the share of answered questions at the German transparency platform abgeordnetenwatch.de where citizens can ask MPs all kinds of questions. Controls are the same as in Table 1 (demographic: female, age, partner, children, university, phd; political: number of terms, minister, leading position). All regressions include a constant. Robust standard errors are reported in parentheses. *, **, *** denote significance at the 10, 5, 1 % level.

Table 4.6 checks the effect of the list positions on absence rates for list MPs in the parliament elected in 2009. Specification (1) includes party controls. Note that this time several more parties are included in the analysis since the proportional tier of the system also allows smaller parties to enter parliament. Specifications (2) and (3) add demographic and political controls on the MP level. Specification (4) adds the percentage of platform answers discussed before, specification

(5) uses unexcused absences only as dependent variable. Overall, there seems to be a significant negative effect of the list position on an MP's absences: Being placed one position further down on the party list (implying a tougher struggle for reelection) reduces an MP's absence rate by between 0.2 and 0.3 percentage points. This preliminary extension thus suggests that also for list MPs some kind of political competition seems to matter for attendance behavior in parliament. Similarly to the analysis of district MPs, there may be endogeneity concerns here as well. However, in contrast to before, the proportional tier of the electoral system does not allow for an IV strategy since party conventions putting together party lists are actually black boxes from the perspective of the researcher. Thus, further research is needed to shed more light on list MPs and the nature of competition they are faced with.

4.8 Conclusion

In this paper, I investigated the relationship between political competition and rent-seeking: Does stiffer electoral competition reduce absences in parliament? – To answer this question, I constructed a novel data set on absence rates of German MPs also including election results and a wide range of MP specific controls. For opposition MPs, there indeed is a strong and significant effect: A 10 percentage points smaller vote margin implies a 7 percentage points smaller absence rate, i.e. MPs under high electoral pressure shirk less. MPs of governing parties are considerably less responsive to their vote margins and are also less absent in general. These results are robust to employing an IV strategy based on using the party vote share as an instrument for an MP's vote margin, to the type of measurement of absences, to controlling for safe positions on party lists, to controlling for an MP's motivation, to a selection story of putting high quality candidates in contestable districts, and to an extension to the parliament elected in 2005. One can also show that MPs with high absence rates are indeed punished by voters in subsequent elections. An extension shows that also MPs elected via party lists respond to stiffer electoral competition by attending more sessions. To my best knowledge, this is the first analysis to use recorded votes in the German parliament to create a measure for parliamentary absenteeism and the first in general to look at list MPs. It is also the first study to look at the relation between political competition and MP attendance behavior in a non-Italian context.

From a normative perspective, it seems worthwhile to enhance electoral competition wherever possible. One way of doing so, suggested by Galasso and Nannicini (2011), may be "optimal gerrymandering", meaning to design constituency borders in a way to minimize the number of districts one of the parties will win for sure in an election.[33] This study supports the view that political rent-seeking might indeed be reduced through adequate electoral reform. Examples for debates on electoral reform from Spain and Britain have been mentioned in the introduction. Also in Germany, redesigning the borders of constituencies is repeatedly being debated.[34] A recent census (the first one since 1987) may likely necessitate redesigning electoral borders anyway because of changed population size numbers per district. In California, an independent commission of citizens (instead of politicians themselves) has now taken control over drawing electoral boundaries to make elections more "competitive" since "in the 612 races of California's last four elections only seven seats have changed from one party to another".[35]

[33] Besley and Preston (2007) illustrate how electoral districting that is biased towards one party can lead to more extreme policy choices. Coate and Knight (2007) analyze potential welfare gains from optimal districting. Both papers, however, focus on a purely majoritarian system.

[34] See, for example, Handelsblatt (2011).

[35] See Economist (2011c).

Chapter 5

Is Status Quo Bias Explained by Anchoring? Evidence from Survey Experiments[1]

5.1 Introduction

In a seminal paper, Samuelson and Zeckhauser (1988) used several survey experiments and field studies to show that people stick to the status quo choice or policy more frequently than predicted by standard economic theory. Everyday examples are returning to the same vacation spot each year or the incumbency advantage in politics. One example investigated by Samuelson and Zeckhauser (1988) is the following: In a survey experiment, they put respondents into the hypothetical role of a commissioner distributing water between a town and farmers during a dry period by choosing one out of ten different options where one of the options was framed as status quo. They found that in treatments where a more farmers-friendly distribution was framed as the status quo respondents actually chose a more farmers-friendly option; with a town-friendly status quo respondents chose a town-friendly option. They labelled this behavior "status quo bias" and outlined several possible explaining theories for it – ranging from rational ones based on economic reasoning such as transition cost to cognitive misperceptions such as anchoring.

[1]For helpful comments and discussions, I thank Gerard van den Berg, Christine Binzel, Pierre Boyer, Dennis Boywitt, Richard T. Carson, Dirk Engelmann, Jana Friedrichsen, Hans Peter Grüner, Zohal Hessami, Dan Houser, Eckhard Janeba, Christian Koch, Tom Krebs, Florian Misch, Jörg Oechssler, Henrik Orzen, Alexander Paul, Stefan Penczynski, Charles R. Plott, Alex Roomets, Christoph Vanberg, Roland Vaubel, Johannes Voget, Klaus Wälde, Evguenia Winschel, and Philipp Zahn. I further thank seminar participants at the Mainz International Behavioral Economics Workshop, at the European Conference of the Economic Science Association, at the Heidelberg Mannheim Experimental Economics Workshop, and at the University of Mannheim, the teams of GIP and GESIS, and the many pretesters of the surveys.

However, their paper remains silent about how to disentangle the different explanations. The purpose of this paper is to use survey experiments to improve our understanding of *why* people overly stick to the status quo and to empirically disentangle different explanations of status quo bias. More specifically, I aim to show how quantitatively important anchoring is as an explanation.

Anchoring describes the phenomenon that people's judgments are biased towards initially presented values. Tversky and Kahneman (1974) were among the first to analyze anchoring. They had people turn a wheel of fortune with numbers from 0 to 100 that was manipulated to show only the numbers 10 or 65 as result (as an "anchor") and asked them afterwards to estimate the number of African countries in the United Nations. People in the anchor 10 group gave a median estimate of 25; people in the anchor 65 group gave a median estimate of 45. Another very prominent illustration of the anchoring effect was given by Ariely et al. (2003): They asked people for the last two digits of their social security number[2] (as an "anchor") and presented them a bottle of wine afterwards. Then they asked if people would be willing to pay more or less than the just stated (social security number digits) number in dollars for the wine. They also asked for the exact maximum willingness to pay in a follow-up question. Surprisingly, it turns out that there is a very robust positive correlation between the social security number digits and the willingness to pay for the wine.[3] Anchoring might therefore be a very relevant factor contributing to status quo bias, especially when numbers are involved like in several examples presented by Samuelson and Zeckhauser (1988). In short, anchoring is a "cognitive misperception" that is seen by Samuelson and Zeckhauser (1988) as "the best explanation" of status quo bias (p. 41).

This paper aims at disentangling anchoring from economic reasoning causes of status quo bias and also at making a first step towards quantifying the anchoring effect in relation to them. Although status quo bias and anchoring are ubiquitous phenomena to analyze, I restrict attention to survey questions as a natural application in this study. Opinion polls via surveys are one of the main devices in representative democracies to get to know citizens' preferences for wide ranges of topics. Besides, survey data find increasing use in economics (e.g. Alesina et al. (2014)) and surveys are well suited for implementing large-scale experiments (e.g. Cruces et al. (2013)). To my best knowledge, this is the first study to isolate and quantify anchoring from other causes of status quo bias in stated preferences.[4] For that purpose, large survey

[2] The first ones to use social security numbers in a similar context were Chapman and Johnson (1999).

[3] This result has, for example, recently been replicated by Bergman et al. (2010) and contested by Fudenberg et al. (2012) and Maniadis et al. (2014).

[4] Anchoring as such has been quantified before. Early examples are Jacowitz and Kahneman (1995) and Green

experiments asking for respondents' policy preferences have been conducted.

Better understanding the relevance of anchoring in shaping status quo bias is an interesting endeavor for at least three reasons: First, we really need to understand better *why* people overly want to stick to the status quo at a very general level. Is it due to cognitive causes or some form of economic reasoning? – If people typically have good reasons for not wanting to change the status quo or have at least seriously thought about it, then politicians may in the end simply have to accept it. However, if it turns out that people always pick the status quo just because of cognitive anchoring and do not really think about the issue at hand at all, it may in fact be a worthwhile political endeavor to try to convince voters that more political reforms may actually be beneficial. Better understanding the causes of status quo bias can in that sense inform politics and the general public about how to optimally deal with status quo bias in politics and elsewhere.

Second, we need to know a lot more about how status quo bias and anchoring affect survey answers in particular. Billions of dollars are spent on surveys where people are asked for preferences and often to give numbers also (Carson and Groves (2007)). Surveys dominate marketing and influence politics. How seriously should we take such surveys? – In case anchoring and purely cognitive effects dominate respondents' decisions, surveys should be considered a less informative tool compared to a situation where economic and rational reasoning determine respondents' answers. My results can also be used to draw some lessons about future survey design.

Third, regarding the relevance of anchoring in particular, we need to understand better how non-economists react to numbers. Anchoring occurs in a wide range of different situations (see Fehr and Tyran (2008), Beggs and Graddy (2009), or Fujiwara et al. (2013)). This paper is one of the few analyzing anchoring in a representative sample outside of the laboratory or the classroom. If we do not know enough about how people treat numbers, we misinterpret our results. – Suppose it turns out that the size of the anchoring effect is typically negligible, then we may not have to worry too much. But in case one finds that anchoring is a relevant force, one may have to rethink some results in the literature, particularly in experimental economics and any other discipline relying on surveys and experiments. Let me give an example to be more specific. In public good financing experiments presenting people the number of what others contribute on average and subsequently observing that people stick to that number themselves may in fact be due to conformity, i.e. people wanting to follow the reference point set by others.

et al. (1998). But these approaches are not applicable to policy preferences and do also not allow for disentangling anchoring from other effects.

And often such behavior is indeed interpreted as evidence of "conformity" (e.g. Alpizar et al. (2008)). However, if anchoring is relevant, people just take the number they are informed about and do in fact not think about the contribution to the public good. In that case, it would be misleading to call this action resulting from a cognitive misperception "conformity". Thus, learning about how quantitatively important anchoring is can inform us about the validity of parts of the research in experimental economics and about how future research designs should be modified.

The experimental approach implemented in this study relies on a split sample survey design, i.e. different respondents get different versions of the surveys. The survey treatments developed for disentangling anchoring from other potential causes of status quo bias are based on Samuelson and Zeckhauser (1988). I complement their approach by adding an anchoring treatment. This anchoring treatment very much follows the social security number design implemented by Ariely et al. (2003). It allows me to identify the anchoring effect and to quantify its relevance relative to other causes of status quo bias. The surveys have all been conducted in 2012. In total, more than 1,500 respondents participated. The main survey has been conducted online and is representative of the German population. For comparison, a smaller survey has been conducted paper-based in the classroom with only students taking part. The policy topic of application for the preferences surveys were different areas of public spending in Germany. The results suggest that anchoring alone can cause one half of the whole phenomenon referred to as status quo bias. This indeed supports Samuelson and Zeckhauser (1988) who suspect anchoring to play a dominant role in explaining status quo bias. When designing future surveys and experiments, the potential prevalence of strong anchoring effects should therefore definitely be taken into account. It also seems to be a good idea not to blindly trust status quo bias in voters' stated policy preferences too much.

Apart from the literature on status quo bias and anchoring, this paper also relates to stated preferences research. So far, this literature has however focused more on strategic and hypothetical biases and on the willingness to pay versus willingness to accept difference. See, for example, Cummings et al. (1995), Cummings et al. (1997), Cummings and Taylor (1999), List (2001), Plott and Zeiler (2005), or Plott and Zeiler (2007). My study also relates to the fairly recent literature on default effects and nudges. Anchoring may in some instances be the explanation for why nudging actually works. A very prominent example is the strand of literature starting with Madrian and Shea (2001). They exploited a rule change regarding 401(k) saving plans in the US and found that a relevant share of participants starting saving under the new regime kept both the default contribution rate and the default fund allocation, although in fact only very few had chosen this rate and allocation before the rule change (with no default being in place).[5] Anchoring might be a very relevant explanation in this case. Finally, there is also a political economy literature trying to understand on a more aggregate level why people want to stick with the status quo so often. The most famous example probably is Fernandez and Rodrik (1991).

The next section gives a short overview of potential explanations of status quo bias. Section 5.3 then introduces the design and survey instruments used. Results are discussed in Section 5.4, an extension is covered in Section 5.5. Section 5.6 concludes.

5.2 Explanations of Status Quo Bias

Anchoring as the favorite explanation of status quo bias brought forward by Samuelson and Zeckhauser (1988) is discussed in the Introduction. See Furnham and Boo (2011) for a literature review on anchoring. On a deeper psychological level, there are different explanations for anchoring itself as an empirical phenomenon. Two important ones are "insufficient adjustment" and "selective accessibility". The first one hypothesizes that when people are asked to provide a quantitative answer they use an initially given number as a starting point (be it reasonable or not) and adjust it only partially until a more or less reasonable value (in the eyes of the respondent and depending on the circumstances) is reached.[6] The second explanation states that the anchor activates selected knowledge about the number that is to be given or estimated that fits the anchor value and therefore biases responses towards the anchor.[7] See Epley (2004) for an overview regarding these two main explanations of anchoring. Following the approach to behavioral economics outlined by Matthew Rabin, the present study is however not concerned with psychologically explaining anchoring as such, but with investigating to what degree anchoring and other factors can help explaining economic behavior (cf. Rabin (2013), p. 617).

A lot of other possible explanations of status quo bias – apart from anchoring – have been brought forward. Samuelson and Zeckhauser (1988) categorize them into psychological commitment explanations and rational explanations. For nice overviews focusing on

[5] The importance of default settings has also been shown, for example, in Johnson and Goldstein (2003), Thaler and Benartzi (2004) or Messer et al. (2007).

[6] See Epley and Gilovich (2001), Epley and Gilovich (2006) or Epley and Gilovich (2010). In the contingent valuation literature the anchoring effect is usually labelled "starting point bias", see Green et al. (1998) for an example analyzing anchoring in contingent valuation.

[7] See Mussweiler and Strack (2001).

psychological explanations, see Anderson (2003) or Eidelman and Crandall (2009). Among the psychological commitment explanations Samuelson and Zeckhauser (1988) mention inter alia sunk costs, regret avoidance, and the drive for consistency. The sunk cost explanation relies on the theory that many people are inclined to keep the status quo since they already invested a lot of resources into it. The concept of regret avoidance is closely related and describes the theory that people feel worse after a bad outcome they caused through a new action compared to an equally bad outcome that resulted from doing nothing (i.e. keeping the status quo) since they feel more personally responsible in the first case. The drive for consistency theory implies that people stick to their past decisions to minimize cognitive dissonance regarding their self-perception. All these explanations of psychological commitment in some way or another rely on people already having made a decision in the past. They are "turned off" in my survey experiments by choosing policy preferences as topic. Arguably, in politics no single individual is solely responsible for past or present decisions. If we find status quo bias being present in the area of policy preferences, channels working through psychological commitment should therefore not be relevant.[8] Samuelson and Zeckhauser (1988) mention loss aversion as another possible explanation rooted in psychology. The idea is that the status quo serves as a reference point with possible losses from deviating from it looming larger than possible gains (cf. Tversky and Kahneman (1991) or Hartman et al. (1991)). But even in examples where there is no framing in terms of gains and losses they do find strong evidence for status quo bias and conclude that loss aversion can therefore not be the most relevant explanation. Therefore, also in the policy preferences survey used for this paper, there is no framing in terms of gains and losses.

The second important branch of explanations outlined in Samuelson and Zeckhauser (1988) are the rational ones. They mention transition costs, uncertainty, and costs of search or analysis. Transition costs refer to the simple fact that it might not be possible to change to a new status quo for free. Uncertainty might play a role when people do not have enough information about the possible options and/or are risk-averse. If people are uncertain about the process of changing the status quo itself, trade uncertainty might also be relevant (cf. Engelmann and Hollard (2010)). Costs of searching for and analyzing alternatives to the status quo might be especially relevant when people do not want to invest too much energy in thinking about the decision to be made (cf. Madrian and Shea (2001) or Anderson (2003)). All these explanations are rational in the sense that the individual really thinks about the choice at hand; if the individual sticks to

[8] Regarding policy preferences, an individual may still decide to stick to the status quo since this is what the majority decided. But this is an active decision that requires yet another reason that determines the decision to follow the majority (such as too high cost of analysis on the individual level). Such causes of status quo bias are covered in the following paragraph.

the status quo it does so motivationally and based on some form of economic reasoning. This is in clear contrast to anchoring as a possible psychological cause of status quo bias which is better described as a cognitive misperception. A worthwhile first step towards disentangling different causes of status quo bias therefore is to separate anchoring as a purely cognitive explanation from the remaining ones which can be rooted in economic reasoning.

5.3 Method

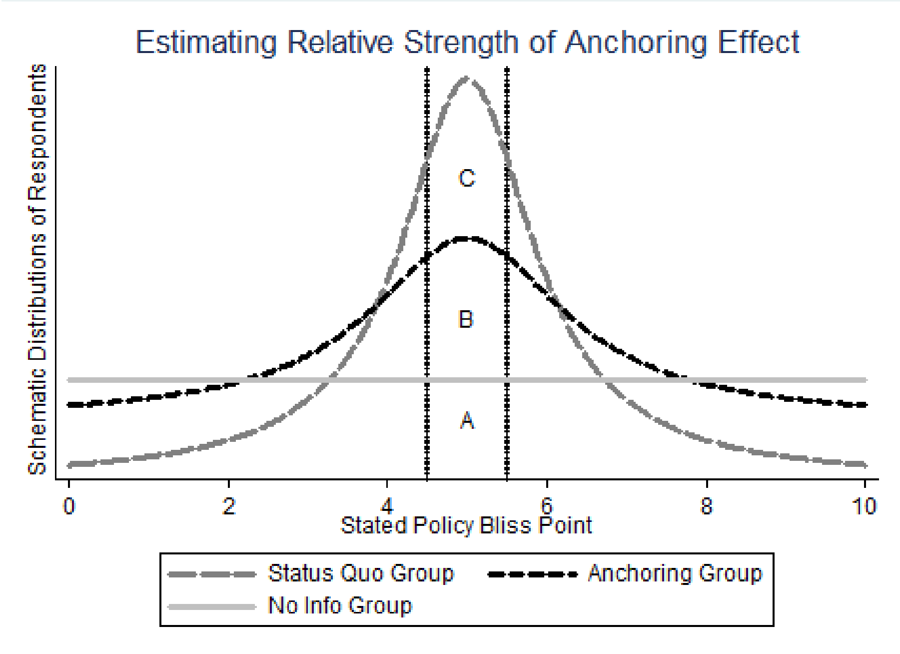
This section first describes the experimental design, then its application to survey questions, and finally the surveys implemented for this study.

5.3.1 Design

A simplified model on how to estimate the anchoring effect's share in contributing to the overall status quo bias effect is depicted in Figure 5.1. The figure shows hypothetical distributions of people's statements across three different experimental treatment groups. The example application is a policy survey asking for people's bliss points in public per capita spending for environmental protection. A split sample design needs to be used, i.e. each respondent gets to answer only one of the three different survey versions. Let us suppose that respondents in the first group, the No Info Group, are asked for their policy bliss point without being given any information on the level of the policy status quo. The solid horizontal line represents the hypothesized distribution of stated policy bliss points within this group. For simplicity, let us assume it is uniform.

The second group, the Status Quo Group, is asked for bliss points for the very same policy as the No Info Group, but also gets informed about the status quo level of the policy. Let us assume it is equal to 5. If status quo bias is an empirically relevant phenomenon, one should expect the distribution of policy bliss points in this group to be more concentrated around 5 than in the No Info Group. This is what the line consisting of long dashes and having a maximum at 5 schematically depicts: A lot more respondents give policy bliss points close to the status quo in the Status Quo Group compared to the No Info Group. One can even give a rough estimate of the status quo bias effect if one agrees on what "close to the status quo" in fact means. For example, let us suppose that everything in between 4.5 and 5.5 is considered close to the status quo. The range of policy bliss points close to the status quo is then marked by vertical dotted lines. Thus, area A represents the number of respondents in the No Info Group that give policy

Figure 5.1: Estimating Relative Strength of Anchoring



bliss points close to 5. For the Status Quo Group, in contrast, respondents close to the status quo are covered by the area $A + B + C$. The absolute status quo bias effect thus amounts to $B + C$, i.e. the difference by which the number of respondents close to the status quo increases when switching from the No Info to the Status Quo Group.

Several different causes may be at work jointly in contributing to the overall status quo bias. The present study is concerned with identifying the relevance of the anchoring effect relative to other causes. Let us now suppose that we can construct a design for a third treatment group, the Anchoring Group, where respondents are asked for their policy bliss points as before, but where by design only anchoring can result in respondents overly stating policy bliss points close to 5. If anchoring is indeed causal for a part, but not all of the whole status quo bias, the resulting distribution could look similar to the one depicted by short dashes in the figure, i.e. being concentrated around 5 as well, but less so than in the Status Quo Group. The part of status quo bias caused by anchoring would be represented by area B. Area B being seen relative to C would give an indication of how relatively relevant anchoring is in explaining status quo bias.

However, one would go too far by saying that $B/(B + C)$ exactly represents the share of status quo bias that is caused by anchoring. For that statement to be true, one would need a linearity assumption stating that anchoring does not interact with any other causes of status quo bias that might be at work in the Status Quo Group besides anchoring, but instead simply adds up to their effects. This is an assumption I am unable to prove and unwilling to make. What can be said, however, is that rational or economic reasoning can by design not account for bunching at 5 in the Anchoring Group and, thus, that anchoring as a purely cognitive misperception can alone produce a $B/(B+C)$ share of the whole status quo bias effect. This gives quite a good first idea on whether anchoring is a relevant factor in explaining status quo bias or not.

The next subsection details how treatments for the three different groups can be designed for the case of policy preferences surveys. Status quo bias is not restricted to survey answers. It is relevant in a wide range of situations where humans make decisions in the presence of a status quo. Nevertheless, it is constructive to follow the example of Samuelson and Zeckhauser (1988) and to use survey experiments as a controlled environment to learn more about this effect. That does not imply that I expect the relevance of my results to be restricted to surveys.

5.3.2 Application to Survey Questions

For illustration purposes, let us stick for the moment to the environmental protection spending example. I come back to the different topics covered in my survey experiments in the next subsection. Screenshots of the original survey instruments that were used in the field are contained in the Appendix.

Using a split sample design, one needs survey treatments for respondents in the No Info Group, the Status Quo Group, and the Anchoring Group. The No Info Group treatment just has to introduce the topic and to ask for the respondent's policy bliss point:

Question A: The German state has taken several measures to protect the environment. In your opinion, how much money should the German state spend on such measures per capita and year?

The treatment for the Status Quo Group additionally contains information on the status quo:

Question A: The German state has taken several measures to protect the environment. It spends around 41 Euros per capita and year on such measures. In your opinion, should the German state spend on such measures per capita and year less than 41 Euros, 41 Euros, or more than 41 Euros?

In the online surveys, the order of the three choice options in Question A is randomized. If a respondent in the Status Quo Group chooses the "more" or the "less" option, she is asked for her exact bliss point in a follow-up question:[9]

Question B: How much money should the German state spend on such measures per capita and year?

Note that the Status Quo Group survey treatment looks similar to the original water commissioner survey question by Samuelson and Zeckhauser (1988) presented in the Introduction. The main difference is that 41 Euros is the actual status quo in German environmental protection spending and that this question could easily be asked in a real policy preferences survey in Germany. Such surveys asking for people's bliss points in certain policy domains while also telling them about the current status quo are very common. Typically they look a lot like the question presented here and could for example also ask for people's preferred marginal income tax rate for the very rich, the average retirement age in a society, or the monetary benefit level of a typical welfare recipient. Note that all survey questions refer to the German case since all the survey experiments have been conducted in Germany.

If status quo bias is relevant, more respondents will give answers close to the status quo in the Status Quo Group version of the survey than in the No Info Group version presented before, although both versions actually ask for the very same thing. This is due to different effects adding up or interacting in more complicated ways to the whole phenomenon labeled status quo bias. The aim is to disentangle the power of anchoring as a cognitive misperception from the remaining other effects contributing to status quo bias such as economic reasoning. When answering the Status Quo Group survey version, respondents are presumably influenced by the number 41 since it represents the status quo policy that is in place (delivering relevant economic information), but also since it is a number (which can serve as a cognitive anchor). If respondents state that their preferred spending level is 41 Euros, we do not know if they want the current policy to stay in place (for possibly motivational reasons based on economic thinking) or if they simply anchored to the 41 as a number for purely cognitive reasons.

Comparisons between the answers to the No Info Group and the Status Quo Group enables one to measure the whole effect of status quo bias. To get an estimate of what part of this whole effect anchoring alone can produce, the Anchoring Group answers a third version of the survey. This anchoring treatment needs to present respondents with a random number to anchor to without telling them anything about the status quo (to "turn off" any other channels through

[9]An alternative would be to skip the choice question and to directly ask for the exact amount. However, real life survey questions typically have the choice question in between. Besides, the anchoring literature from which I borrow the anchoring treatment also uses a choice question. I therefore follow their formulation.

which status quo bias may work as well). If many respondents in the Anchoring Group stick close to the (now random) number in the survey question, this must be due to anchoring. For this purpose, I keep the structure of the question similar to the versions before, but adopt at the same time the anchoring treatment of Ariely et al. (2003). They use people's social security numbers as random anchors. Since people in Germany do not know their social security number by heart, I use phone numbers instead:

Question A: Please enter the last two digits of your phone number (landline, alternatively mobile) into the box provided below.

In the following question, the computer replaces [XX] by the digits the respondent provided in Question A.

Question B: You gave number [XX] as response to the previous question. The German state has taken several measures to protect the environment. In your opinion, should the German state spend on such measures per capita and year less than [XX], [XX], or more than [XX] Euros?

In the online surveys, the order of the three choice options in Question B is randomized. If a respondent in the Anchoring Group chooses the "more" or the "less" option, she is asked for her exact bliss point in a follow-up question:

Question C: How much money should the German state spend on such measures per capita and year?

If many respondents' policy bliss point in the Anchoring Group is close to their phone number, this must be due to anchoring to the number. As explained in Figure 5.1, comparing the share of respondents sticking close to the number between Anchoring and Status Quo Group then allows to get an idea about how relevant anchoring is in contributing to status quo bias.

Overall, the design of the three versions of the survey thus closely follows the original status quo bias analysis format by Samuelson and Zeckhauser (1988), the anchoring analysis format by Ariely et al. (2003), and the way real life policy survey questions typically look like.

One potential concern with my survey design for the Status Quo and the Anchoring Group could be that a lot of respondent simply pick the (exactly) "41 Euros" or the (exactly) "[XX] Euros" because they seem to be "easy answers". However, there are four arguments against this potential concern. First, respondents may not anticipate that they will be asked for the exact amount in a follow-up question in case they pick the "more" or "less" option. In that case all three options are "equally easy". Second, this study is interested in comparisons across different survey treatments. If the just described potential problem occurs to the same degree in both the Status Quo and the Anchoring Group it poses no problem since the effects “cancel out” via the comparison. Third, even if one assumes the problem to be stronger in the anchoring treatment, e.g. because of its non-everyday structure, Epley and Gilovich (2001) have shown that this type of problem is weakest in the case of anchors that are generated by respondents themselves (which is the case with phone numbers respondents give themselves). Fourth, and most important, results are similar when excluding all respondents who took the “41 Euros” in the Status Quo or the “[XX] Euros” in the Anchoring Group.

5.3.3 The Surveys

The experimental design described in the previous subsection was implemented in three different surveys. Table 5.1 gives a short overview. All surveys in principle followed the same basic structure with No Info Group, Status Quo Group, and Anchoring Group. In some cases, there were some additional treatments and checks included. I come back to this in the Extension Section. For simplification, for the most part of the Results Section the focus is on data from Survey 3 since results are very similar across all three surveys.

Table 5.1: Overview of Surveys

	Survey 1	Survey 2	Survey 3
Field Time	Jun/Jul 2012	Sep 2012	Nov 2012
Field Mode	GIP (online)	Classroom (offline)	GIP (online)
Observations	296	176	1081
Policy Topic	Environment	Performing Arts	Performing Arts
Status Quo	41	33	33

Notes: “Status Quo” refers to status quo per capita spending in that area by the German state.

Two out of the three surveys were conducted online using the German Internet Panel (GIP) based at the University of Mannheim and funded by the German Science Foundation. One of its main goals is “to collect micro data on the feasibility of political and economic reforms”.[10] The GIP is anonymous and the first online true probability sample in Germany. Although GIP surveys are conducted online, the recruiting is done offline and is representative. Participants without computer and/or internet access (or lacking the relevant experience) are provided with infrastructure and help. There is a payment of 4 Euros per completed survey, plus a bonus for completing all 5 to 6 surveys in a year. Since the anchoring studies mentioned above (including

[10]For more information on the GIP, you can visit reforms.uni-mannheim.de (choose “Internet Panel” in the menu).

Ariely et al. (2003)) took place either in the classroom or in the laboratory, my study can on top of its original purpose also be seen as the first "field test" of this kind of anchoring in the sense that the subject pool is a representative sample of an entire population.

For comparison reasons, one of the three surveys has been conducted paper-based in the classroom. The sample consisted of second-year undergraduate students taking intermediate microeconomics exercise sessions (in German) at the University of Mannheim. Students were told that the survey was part of a "scientific project", that participation was voluntary and anonymous, and that the results of the project would be presented in one future exercise session. All the students participated and filled out the questionnaire. Students were not aware of the fact that there were different versions of the questionnaire and had also been asked to work on the questionnaire alone. Any communication was forbidden while filling out the sheet of paper. The filling-out process was monitored. The classroom survey allows checking whether results regarding status quo bias and anchoring are different for this very specific sample of respondents compared to the population-representative surveys. Respondents in the classroom survey have all passed the highest possible high school degree in Germany (qualifying for entering university), are on average considerably younger, and have on average a background of two semesters training in economics and mathematics at the University of Mannheim.

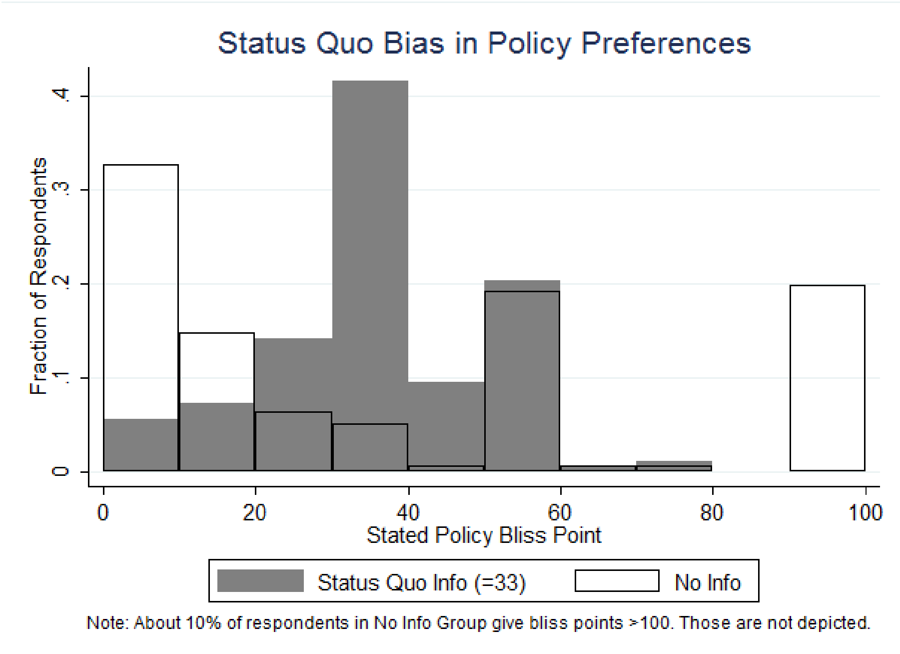
5.4 Results

This section first presents evidence for status quo bias in people's policy preferences. The second part shows that people are prone to anchoring to random numbers when stating policy preferences. The third part analyzes how quantitatively relevant anchoring is in contributing to status quo bias.

5.4.1 Evidence of Status Quo Bias

Figure 5.2 comparatively shows the distribution of respondents' stated policy bliss points in the No Info and the Status Quo Group of Survey 3. Although both groups are in fact asked for the very same thing (their policy bliss point regarding public per capita spending in Euros on the performing arts in Germany), informing one group about the status quo (but not the other) results in large differences across the distributions of answers: One can immediately tell that in the No Info Group answers around 0, 50, and 100 seem prominent magnitudes to pick, whereas in the Status Quo Group the most striking observation clearly is that as much as 40% of all respondents give a response equal to 33 or at least very close to it. They had been informed that 33 is the status quo policy. Thus, respondents are heavily influenced by knowing about the status quo or not. Status quo bias is strong.

Figure 5.2: Status Quo Bias in Policy Preferences



It is also evident from the figure that answers in the No Info Group are more dispersed than in the Status Quo Group. This observation further supports the conclusion that status quo bias is at work and is confirmed by looking at some summary statistics of the bliss point distributions across the two groups. For this, see Table 5.2. While the range of given answers is 0 to 699 in the No Info Group, it shrinks to 0 to only 75 in the Status Quo Group. The standard deviation of answers is more than seven times higher in the No Info Group than in the Status Quo Group. Finally, the mean policy bliss point is around 64 in the No Info Group, but exactly equal to the status quo, i.e. 33, in the Status Quo Group. Thus, informing respondents about the status quo has enormous effects on the answers they give. The clearest evidence of status quo bias is that almost one half of all respondents give policy bliss points equal or very close to the status quo if they get informed about it. Results are very similar for the other surveys.

Table 5.2: Status Quo Bias in Policy Preferences

	No Info Group	Status Quo Group
Mean	63.76	33.08
Stdev	98.55	13.81
Min	0	0
Max	699	75
Obs	177	178

Notes: Data from Survey 3. No info group did not receive any information on current status quo policy. Status quo group was informed that current status quo policy is 33. The difference in means across the two groups is statistically significant (p=0.0000). The total number of observations may seem low compared to Table 5.1. Note, however, that not all treatments implemented in Survey 3 are already depicted in Table 5.2.

5.4.2 Evidence of Anchoring

Part of the status quo bias found in the previous subsection may be due to cognitive anchoring to the number 33 presented to respondents in the Status Quo Group. To investigate this further, this subsection presents evidence for anchoring to numbers from the Anchoring Group in Survey 3. In the Anchoring Group, respondents are asked to relate their policy bliss point to their phone number statement. See Table 5.3 for results from the Anchoring Group. It depicts respondents' stated policy bliss points for different quintiles of the distribution of phone number digits stated by the respondents right before the bliss point. The results suggest that respondents who give a high phone number also give a high policy bliss point: The mean bliss point almost monotonically increases with the quintile of the phone distribution. While the 20% of respondents with the lowest phone numbers give a mean policy bliss point of 22, the 20% of respondents with the highest phone numbers give a mean policy bliss point of 72. The correlation between phone digits and bliss points is 39% and highly significant.

Even when excluding all respondents who took neither the "*more than [XX] Euros*" nor the "*less than [XX] Euros*", but the "*[XX] Euros*" option, the anchoring correlation is still 33% and highly significant (p=0.0000). Results are very similar for the other surveys. For example, also in Survey 2 where all respondents were highly educated university students having been trained in economics and mathematics the phone anchoring correlation is 37% and significant at any conventional level.

Thus, we can conclude: People with a high phone number (high anchor) want high per capita spending on the arts, while people with a low phone number (low anchor) want low spending. People take their own phone number as a relevant anchor when answering the policy question.

Table 5.3: Anchoring to Own Phone Number

Quintile of Phone No.	Mean Policy Bliss Point
First	22
Second	37
Third	36
Fourth	56
Fifth	72
Correlation	0.39*** (p=0.0000)

Notes: Data from Survey 3. 345 observations. The first column is based on respondents' own announcing of phone numbers. Column two gives mean stated policy bliss points for the respective quintiles of the phone number distribution. The correlation refers to the correlation between the stated phone numbers and the stated policy bliss points.

Anchoring is very relevant. This finding is completely in line with the results of Ariely et al. (2003) who show that people with a larger social security number state a higher willingness to pay for the same bottle of wine.

The next subsection details how the experimental design can be exploited to investigate how much of the status quo bias detected before is due to such anchoring.

5.4.3 Explaining Status Quo Bias by Anchoring

A first step in analyzing how quantitatively important anchoring is in causing status quo bias is to compare between the Status Quo and the Anchoring Group the shares of respondents giving policy bliss points within a certain "closeness" band around the status quo or anchor. In both groups, respondents get a number to anchor to, be it the status quo or a phone number. In the Status Quo Group, respondents may besides anchoring to the number also give a policy bliss point close the status quo number for other potential reasons of status quo bias (such as based on economic reasoning). If one accepts the share of respondents staying close to the status quo as a measure of the total status quo bias effect, one is able to estimate the share of this effect that is due to anchoring by comparing it to the share of respondents staying close to the phone number in the Anchoring Group (where all other channels possibly causing status quo bias are turned off).

Since it is not a priori clear what "staying close" means, Table 5.4 provides such an analysis for bands of different ranges from +/-1% up to +/-50%. For example, being in a 1% band means that the answer given to the question was at maximum 1% larger and at minimum 1% smaller than the number given in the question (be it the status quo policy or the phone anchor). One

Table 5.4: Anchoring Can Cause up to 50% of Status Quo Bias

Band around anchor/status quo	Respondents fraction within band: No Info	Anchoring	Status Quo	Chi square p-value	Relative strength anchoring effect
1%	0.000	0.135	0.376	0.000	0.359
5%	0.000	0.141	0.376	0.000	0.375
10%	0.045	0.156	0.416	0.000	0.300
20%	0.045	0.207	0.416	0.000	0.437
30%	0.056	0.305	0.556	0.000	0.498
50%	0.107	0.443	0.652	0.000	0.617

Notes: Data from Survey 3. 689 observations (sum of observations analyzed separately in Tables 5.2 and 5.3). Column one gives the +/- range by which a respondent's stated policy bliss point can deviate from the anchor (Anchoring Group) respectively status quo (No Info and Status Quo Groups) to still be considered close. For example, 13.5% of respondents in the Anchoring Group stay within a +/- 1% range of the anchor. Column four gives the p-value of a chi square test checking whether the fraction of respondents within the band is equal across the three groups. The last column gives a rough estimate of the strength of the anchoring effect relative to the status quo bias by calculating: (fraction anchoring - fraction no info) / (fraction status quo - fraction non info).

can see that in the Anchoring Group between 13.5 and 44.3% of all respondents (depending on the definition of "staying close") give a policy bliss point that is close to their phone number statement. In the Status Quo Group, between 37.6 and 65.2% of respondents give a policy bliss point that is close to the status quo of 33. Table 5.4 also presents p-values of chi square tests showing that the shares of respondents sticking close to the number are indeed statistically different across the different treatment groups. The final column of Table 5.4 – in line with the design presented in Figure 5.1 – provides back of the envelope estimates of the relative importance of the anchoring effect in contributing to the overall status quo bias effect. They show that according to Survey 3 data cognitive anchoring alone can produce between 30 and 61.7% of the whole effect referred to as status quo bias. Analyzing the data from the other surveys gives the same picture. See the Appendix for the results.

Table 5.5 offers a more formal analysis by regressing the respondents' distances between policy bliss point and number (status quo or anchor) on treatment group indicators. The No Info Group serves as a baseline, i.e. the mean distance between the status quo of 33 and policy bliss points stated by respondents in the No Info Group is about 56.2. The coefficients for both other groups are significantly negative, i.e. providing respondents with the actual status quo or asking them to relate the bliss point to their phone number both pushes them closer to the number. Being informed about the status quo policy reduces the distance to the status quo by about 46.6 to a mean distance of about $56.2 - 46.6 = 9.6$ only. This is again illustrative of status quo bias. Being a respondent in the Anchoring Group reduces the distance from the relevant number by only 27 resulting in a mean distance of $56.2 - 27 = 29.2$. Comparing the coefficients

Table 5.5: Regressing Distance from Anchor/Status Quo on Group Indicator

Dep. Variable:	Distance from Anchor/Status Quo
Constant	56.198 ***
	(6.500)
Status Quo	-46.552 ***
	(6.541)
Anchoring	-26.979 ***
	(6.709)
Observations	689
R squared	0.106

Notes: Data from Survey 3. OLS regression. 689 observations (sum of observations analyzed separately in Tables 5.2 and 5.3). Dependent variable is the absolute distance of the respondent's stated policy bliss point from the anchor (anchoring group) or the status quo (no info and status quo group). Status Quo and Anchoring are dummies indicating to which group a respondent had been randomly assigned to. The no info group serves as the baseline. The coefficients of the Status Quo dummy and the Anchoring dummy are significantly different (p=0.0000). Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

of the treatment group indicators suggests that about $27/46.2 = 58.4\%$ of status quo bias can be produced by anchoring alone based on this distance analysis. This magnitude is very much in line with the estimates obtained in Table 5.4 based on an analysis of shares of respondents within certain closeness bands.

Thus, the estimates show across the different surveys and across different methods of analysis that anchoring alone can produce about one half of the whole effect referred to as status quo bias (between 30 and 62% according to Table 5.4, 58% according to Table 5.5). Although the results are fairly consistent across the survey experiments, one may question the external validity of the result since all presented survey experiments are based on asking common people about detailed policy issues such as per capita spending on environmental protection. One could argue anchoring to be particularly strong in this case since many people may not have the slightest clue about what a potential range of reasonable answers could be. I want to alleviate such concerns by bringing up three points: First, the used survey questions follow very much real life survey questions. Thus, even if they were particularly prone to anchoring by non-experts, so would be real life surveys. In that sense, my results would still be informative for real life applications. Second, Bateman et al. (1997) have shown that people show about the same degree of uncertainty regarding preferences in the case of more common goods compared to less common goods. This fits the findings by Jacowitz and Kahneman (1995) and Fujiwara et al. (2013) showing that experts are in fact not less prone to anchoring than non-experts. Third, in some of the experimental surveys, I can control for the experience of the respondents, e.g. by asking respondents how much money they typically spend per year on visiting stage performances before asking them about their bliss point regarding public spending on the performing arts. My results are similar when controlling for such experience.

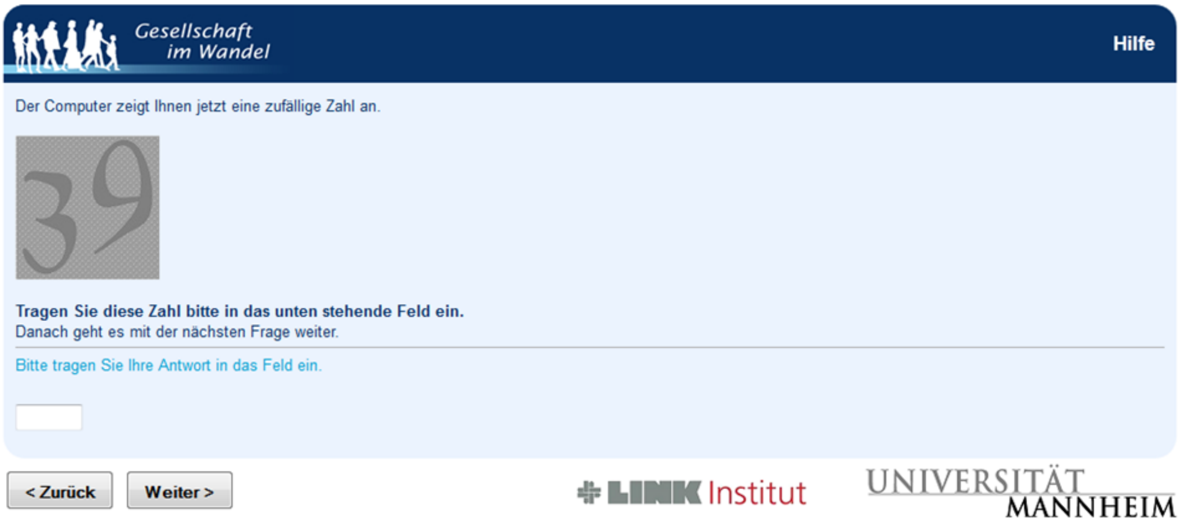
5.5 Extension: The Limits of Anchoring

The results in the previous section show that anchoring as a cognitive misperception can explain up to 62% of status quo bias. If anchoring has such strong effects, the question arises whether policy preferences may also be influenced by any random number even if it is not connected to the policy preferences question at all. In the previous design, respondents in the Anchoring Group were explicitly asked to relate their policy bliss points to their phone number statements. This subsection explores an extension treatment to the previous design that tests whether respondents also anchor when they are presented a random number on the way without being actively asked to consider it when answering the subsequent policy preferences question. The psychological literature refers to this special type of anchoring as “basic anchoring”. Results on basic anchoring have been mixed so far and seem to depend on the circumstances. While Wilson et al. (1996), Mussweiler and Englich (2005), and Critcher and Gilovich (2008) find basic anchoring effects, Brewer and Chapman (2002), for example, do not.

Figure 5.3: A Typical Captcha

To design a basic anchoring treatment, one needs to present respondents a number (to potentially anchor to) without giving them the slightest hint that there could be a relation to the subsequent policy preferences question. This turns out not to be too straightforward if one does not want to lie to people and also not to make them suspicious. I decided to use a "captcha" for that purpose. Captchas are "Completely Automated Public Turing tests to tell Computers and Humans Apart" and known to everybody using the internet (see von Ahn et al. (2003), von Ahn et al. (2008), or Rao and Reiley (2012)). Figure 5.3 shows you a typical example.[11] Robots or programs are often used to misuse websites, but are often unable to complete captchas. Captchas are therefore used as a security device deciding whether the entity using a website is human or not. Since such captchas are by now very common, I can use a captcha respondents have to answer in a question right before the policy question to potentially get an anchor into respondents' minds without creating suspicion. See Figure 5.4 for a captcha adapted for my purposes. Respondents are asked to enter the random number the computer shows them into the box provided.

Figure 5.4: An "Anchor 39" Captcha



To test whether including such basic anchors has any effects on respondents' stated policy bliss points, Survey 3 contained an additional treatment group (Basic Anchoring Group). Respondents in this group first had to answer a captcha question like the one shown in Figure 5.4. The computer randomly showed either an anchor 33, an anchor 66, or an anchor 99 captcha. Afterwards, they were presented with exactly the same policy preferences question as respondents in the No Info Group. See the Appendix for detailed screenshots and explanations of the original survey questions. The results in Table 5.6 show that respondents in the Basic Anchoring Group

[11]The example is taken from the following website that explains how to create a captcha: http://dotnetslackers.com/articles/aspnet/Creating-a-CAPTCHA-generating-ActionResult.aspx (March 18, 2014).

do in fact not anchor to the captcha numbers at all: Independently of the captcha number being shown to them, in all three groups respondents' stated mean policy bliss points are almost the same. There is no evidence of any correlation between the stated policy bliss points and the captcha numbers. Respondents do thus not exhibit basic anchoring.

Table 5.6: No Anchoring to Captchas

Captcha	Mean Policy Bliss Point
33	72.85
66	70.43
99	69.62
Correlation	-0.01 (p=0.8361)

Notes: Data from Survey 3. 349 observations. Column one presents the captcha number to be filled in by respondents in the respective group. Column two gives the mean stated policy bliss points for each captcha group. The correlation refers to the correlation between captcha numbers and stated policy bliss points.

Since real life survey questions do in fact actively ask for a comparison of the number given in the question (e.g. the status quo) to the respondent's bliss point, the Anchoring Group design (with phone numbers) is the appropriate one to estimate anchoring's relative effect in contributing to status quo bias. The previous section has shown that this relative effect is strong. Nevertheless, it may seem assuring that results from the Basic Anchoring Group show that, although anchoring can be strong, people are not necessarily prone to taking any random number not related to the question at hand at all as relevant information when answering it.

5.6 Conclusion

Status quo bias is a ubiquitous phenomenon. However, little is known about which factors can explain it. This study uses large survey experiments within the classroom and the representative German Internet Panel (GIP) to disentangle anchoring as a cognitive misperception from explanations based on economic reasoning such as transition cost. The result is that about one half of status quo bias can be caused by anchoring alone. This clearly supports the claim by Samuelson and Zeckhauser (1988) that anchoring is the "best explanation" of status quo bias. One lesson to be learnt from that could be not to take huge status quo bias in voters' stated policy preferences politically too seriously. Another implication is that anchoring as an empirical phenomenon should be taken into account when designing future experiments and surveys, in particular when presenting people with numbers. An extension analysis relying on the usage of captchas as basic anchors however reveals that fears people could systematically anchor to any random number that has no connection to the issue at hand at all would be exaggerated.

Appendix A

Appendix to Chapter 2

A.1 Maps

The following two maps (Figure A.1 and Figure A.2) show the share of years between 1978 and 2010 in which a state reformed the welfare system and in which a state had a divided government. The darker the color the stronger the incidence.

Figure A.1: Reform Incidence Across States

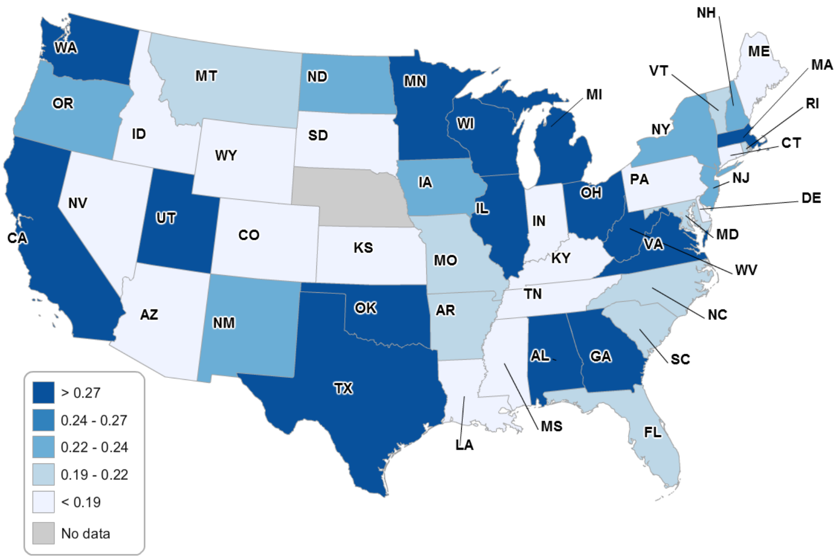
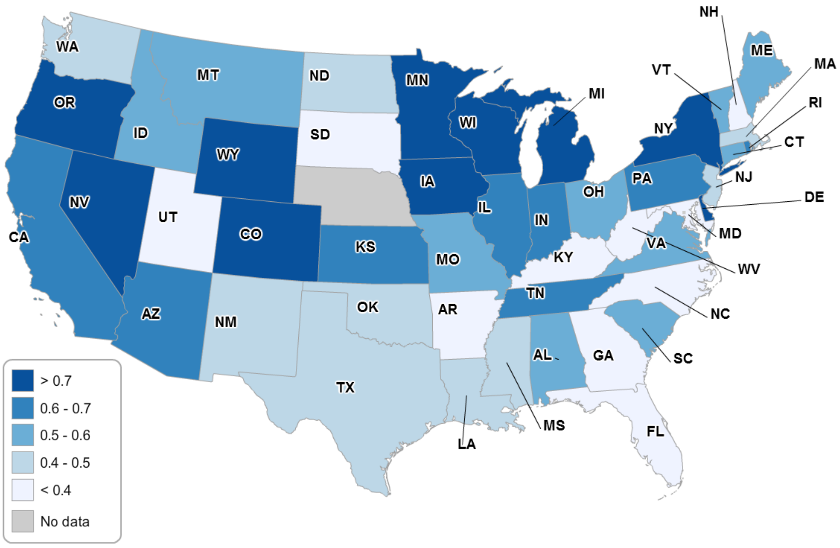


Figure A.2: Divided Government Incidence Across States



A.2 Differential Effects in Different Samples

Electoral Cycles and Government History. This section explores the robustness of the result across different sample restrictions. The first restriction is using electoral cycles as units of observation instead of years. This roughly cuts the sample into half since usually every other year there is at least a legislative election in most states. The reason to look at electoral cycles as a robustness check is that under divided government reform-making may be staggered across several years leading to a correlation between divided government and reform years. The mechanism could be as follows: A unified government can easily implement a welfare reform in its first year in office since it is not confronted with any institutional obstacles, while a divided government has to struggle more, gets blocked, and consequently implements only a part of the reform in the first year and a second part in the second year in office. Since the unit of observation in the baseline analysis is state-years, the reform dummy would be equal to one for the first year for the unified government, but would be equal to one for both the first and the second year for the divided government. Although the two governments may in fact have implemented the very same reform, the differences in timing may make the divided government look more reformist in the analysis.

This is what Appendix Table A.1 investigates by looking at whole electoral cycles instead of years, i.e. the reform dummy is now defined on the level of governments and no longer on a yearly basis. Specification (1) shows that the effect of divided government is still sizable and significant. The finding is thus not driven by differences in reform timing and the mechanism outlined in the preceding paragraph cannot account for it. Appendix Table A.1 also checks on the level of governments if the divided government effect is differential depending on the previous government. One could, for example, imagine a divided government to be more likely to reform when it follows on a unified Democratic government which tried to keep the status quo. Or a divided government could be more likely to reform when the previous government also had been divided and the party leaders are already used to the situation of non-unified partisanship. Specifications (2) to (4) reveal that no such differential effects can be identified. Thus, it seems that divided government per se has an enhancing effect on reform adoption – irrespectively of the past form of government. Specification (5) looks at how the effect may be different depending on how a divided government came to power – via general elections (when the governor is up for election) or via midterm elections (when the governor is not up for election). About one third of the changes from unified to divided government are the result of midterm elections, while two thirds result from general elections. As can be seen from specification (5), the estimated coefficients' sizes are different, but not statistically significant. One can therefore not conclude that the divided government effect was solely driven by general or by midterm elections.

Table A.1: Divided Government, Government History and Reform

	(1)	(2)	(3)	(4)	(5)
Divided Government	0.0899**				
	(0.0406)				
Divided Govt. with past Unified Govt.		0.0834			
		(0.0603)			
Divided Govt. with past Divided Govt.		0.0698		0.0697	
		(0.0507)		(0.0510)	
Divided Govt. with past Unified Dem. Govt.			0.0611	0.0883	
			(0.0813)	(0.0812)	
Divided Govt. with past Unified Rep. Govt.			0.0409	0.0730	
			(0.0613)	(0.0711)	
New Divided Govt. via General Elections					0.0334
					(0.0772)
New Divided Govt. via Midterm Elections					0.101
					(0.0884)
Demographic Controls (lagged one year)	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES
State FE	YES	YES	YES	YES	YES
State Specific Linear Trend	YES	YES	YES	YES	YES
Observations	651	651	651	651	651
R-squared	0.489	0.487	0.484	0.487	0.485

Notes: The dependent variable in all specifications is a reform dummy that is equal to one if one or more welfare reforms have been introduced in a given state by a given government. Divided Government is a dummy that is equal to one when either the majority of the state's lower legislative chamber or the majority of the state's upper legislative chamber is from another party than the governor. For details regarding any of the variables, see the Data Appendix. Robust standard errors clustered at the state level are shown in parentheses. *** $p<0.01$, ** $p<0.05$ and *$p<0.1$.

Space and Time. Appendix Table A.2 differentiates the effect of divided government on the likelihood of reform adoption across space and time. Specification (1) shows the baseline specification from before. Specification (2) adds an interaction for Southern states. This allows the South of the US – which is commonly known to be potentially politically different compared to the rest of the US – to be different with respect to the effect of divided government on reform. It turns out that this seems not to be important: The coefficient of the interaction term is small and not significant. Specification (3) adds an interaction with a dummy that is equal to one for all years after 1996 when the landmark Welfare Reform under President Clinton was implemented. As outlined before, this reform fundamentally changed welfare politics in the US. The estimation results show that the effect is not statistically different before and after the 1996 Welfare Reform. The positive coefficient of the interaction term seems to suggest that the divided government effect is a bit larger after 1996, but this can be explained by the fact that also the unconditional propensity to reform welfare in a given year is larger after 1996. Moreover, statistical power is not sufficient to identify the divided government effect separately for the two time periods. Specification (4), finally, separates the highly reformist Clinton era (1992-2000) from the Non-Clinton era. Although, the interaction effect for the non-Clinton era is statistically not significant, the negative sign seems to suggest that the divided government effect may have been indeed stronger during the presidency of Clinton.

Table A.2: Divided Government and Reform in Different Samples

	(1)	(2)	(3)	(4)
Divided Government	0.0649***	0.0764**	0.0535	0.105**
	(0.0229)	(0.0315)	(0.0342)	(0.0459)
Divided Government * Southern State		-0.0325		
		(0.0442)		
Divided Government * After 1996 Dummy			0.0248	
			(0.0627)	
Divided Government * Non-Clinton Era				-0.0578
				(0.0457)
Demographic Controls (lagged one year)	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
State FE	YES	YES	YES	YES
State Specific Linear Trend	YES	YES	YES	YES
Observations	1,343	1,343	1,343	1,343
R-squared	0.326	0.326	0.326	0.326

Notes: The dependent variable in all specifications is a reform dummy that is equal to one if one or more welfare reforms have been introduced in a given state and year. Divided Government is a dummy that is equal to one when either the majority of the state's lower legislative chamber or the majority of the state's upper legislative chamber is from another party than the governor. For details regarding any of the variables, see the Data Appendix. Robust standard errors clustered at the state level are shown in parentheses. *** $p<0.01$, ** $p<0.05$ and *$p<0.1$.

A.3 Further Robustness Checks

This section shows some further robustness checks. Appendix Table A.3 illustrates that the baseline results presented in Table 2.2 are robust with respect to esimating logit instead of linear probability models. The estimated odds ratio is about 1.5, meaning that divided governments are about 50% more likely to reform than unified governments.

Appendix Table A.4 shows that the ideology robustness results from Table 2.4 also hold when restricting attention to only contractive welfare reforms. The smaller coefficient in Appendix Table A.4 compared to Table 2.4 can be explained by the fact that also the unconditional reform probability is lower when focusing on contractive reforms only. Expansive reforms are empirically irrelevant before 1996, and even after 1996 they do account for less than 25% of all reforms.

Appendix Table A.5 uses Arellano-Bond estimation to lend further credibility to the robustness results with respect to the reform history presented in Table 2.5.

Appendix Table A.6 shows that the divided government effect is also present when using count variables reform measures and Poisson estimation.

Table A.3: Divided Government and Reform (Logit Estimation)

	(1)	(2)
Divided Government (Odds Ratio)	1.534**	1.597***
	(0.254)	(0.286)
Demographic Controls	NO	YES
Year FE	YES	YES
State FE	YES	YES
Observations	1,474	1,343
Pseudo R-squared	0.400	0.468

Notes: The dependent variable in all specifications is a reform dummy that is equal to one if one or more welfare reforms have been introduced in a given state and year. Divided Government is a dummy that is equal to one when either the majority of the state's lower legislative chamber or the majority of the state's upper legislative chamber is from another party than the governor. For details regarding any of the variables, see the Data Appendix. Standard errors are shown in parentheses. *** p<0.01, ** p<0.05 and *p<0.1.

Table A.4: Divided Government, Ideology and Reform (Narrow Contractive Reform 1978-2010)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable: Narrow Reform Dummy								
Divided Government	0.0408**	0.0408**	0.0403*	0.0439**	0.0476**	0.0395**	0.0555**	0.0862**
	(0.0200)	(0.0201)	(0.0201)	(0.0195)	(0.0215)	(0.0196)	(0.0235)	(0.0387)
% Dem. Votes Presid. Elect.	-1.657						-1.221	
	(3.089)						(3.330)	
Citizens Ideology (/1000)		1.367					2.708	
		(2.078)					(2.883)	
Dem. Seats Senate			-0.0742	-0.444*			-0.410*	
			(0.168)	(0.229)			(0.238)	
Dem. Seats House			0.0367	0.189			0.163	
			(0.208)	(0.240)			(0.250)	
Dem. Seats Senate * Dem. M.				0.197**			0.208**	
				(0.0941)			(0.102)	
Dem. Seats House * Dem. M.				-0.0622			-0.0339	
				(0.0748)			(0.0841)	
Government Ideology (/1000)					-0.0977		-1.213	
					(0.650)		(1.492)	
Gov. Dem. (0 = Rep.)						-0.00529	0.0313	0.0476
						(0.0200)	(0.0463)	(0.0442)
Divided Govt. * Gov. Dem.								-0.0872
								(0.0605)
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
State FE	YES	YES	YES	YES	YES	YES	YES	YES
State Specific Linear Trend	YES	YES	YES	YES	YES	YES	YES	YES
Observations	1,474	1,474	1,474	1,474	1,383	1,474	1,383	1,474
R-squared	0.291	0.291	0.291	0.294	0.295	0.291	0.298	0.292

Notes: The dependent variable in all specifications is a reform dummy based on a narrow definition of contractive welfare reforms that is equal to one if one or more welfare reforms have been introduced in a given state and year. Divided Government is a dummy that is equal to one when either the majority of the state's lower legislative chamber or the majority of the state's upper legislative chamber is from another party than the governor. % Democratic Votes Last Presidential Election is divided by 1000. All sears variables refer to seat shares. Government and Citizens Ideology are both calculated from Berry et al. 1998. For details regarding any of the variables, see the Data Appendix. Robust standard errors clustered at the state level are shown in parentheses. *** p<0.01, ** p<0.05 and *p<0.1.

Table A.5: Divided Government and Reform History (Arellano-Bond Estimation)

	(1)	(2)	(3)
Divided Government	0.0819**	0.0837*	0.0868*
	(0.0403)	(0.0440)	(0.0480)
Reform Dummy (t-1)	0.00306	0.0302	-0.00199
	(0.0612)	(0.0976)	(0.115)
Reform Dummy (t-2)		0.0334	-0.0141
		(0.0764)	(0.0822)
Reform Dummy (t-3)			-0.0713
			(0.0597)
Demographic Controls (lagged one year)	YES	YES	YES
Max. lag of dep. variable used as instrument	4	5	6
Observations	1.342	1.342	1.342
Wald Chi Square Statistic	46.07	52.08	32.24

Notes: The dependent variable in all specifications is a reform dummy that is equal to one if one or more welfare reforms have been introduced in a given state and year. Divided Government is a dummy that is equal to one when either the majority of the state's lower legislative chamber or the majority of the state's upper legislative chamber is from another party than the governor. For details regarding any of the variables, see the Data Appendix. Standard errors clustered at the state level are shown in parentheses. *** $p<0.01$, ** $p<0.05$ and *$p<0.1$.

Table A.6: Divided Government and Count Measures of Reform (Poisson Est.)

	(1)	(2)	(3)	(4)
Dependent Variable:	**Broad Reform Count Var.**		**Narrow Reform Count Var.**	
Divided Government	0.154*	0.133	0.192*	0.246*
	(0.0869)	(0.121)	(0.112)	(0.136)
Demographic Controls (lagged one year)	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
State FE	NO	YES	NO	YES
Observations	1,343	1,343	1,343	1,343

Notes: The dependent variables count the number of welfare reforms that have been introduced in a given state and year. The dependent variable in specifications (1) and (2) is based on a broad set of welfare policy rule changes. The dependent variable in specifications (3) and (4) is based on a broad set of welfare policy rule changes. Divided Government is a dummy that is equal to one when either the majority of the state's lower legislative chamber or the majority of the state's upper legislative chamber is from another party than the governor. For details regarding any of the variables, see the Data Appendix. Poisson estimation is employed. Robust standard errors clustered at the state level are shown in parentheses. *** $p<0.01$, ** $p<0.05$ and *$p<0.1$.

A.4 RDD

The section presents a short RDD analysis of the effect of divided government on reform adoption. Compared to the panel data analysis, the main difference in terms of identification is to exploit deeper knowledge about the selection rule determining treatment. In particular, the RDD uses the fact that treatment (divided versus unified government) changes discontinuously in election results (of governors, state houses, and state senates). Focusing on close elections provides us with quasi-experimental treatment assignment. RDD's "randomized variation is a consequence of agents' inability to precisely control the assignment variable near the known cutoff" (Lee (2008), p. 282), i.e. in this setting the many voters' inability to perfectly manage the joint election result makes the institutional setting "quasi-random". Because of the interplay of three different institutions determining treatment, I have three interdependent assignment variables in this RDD. This is non-standard and I have to adjust the design as explained below. Since this complication causes a need of many observations, this RDD analysis should be considered only complementary to the panel data analysis presented before. The remainder of this RDD section first shortly reviews the political economy RDD literature, then explains in a detailed way how to deal with the interdependent assignment variables in the divided government setting, then covers the identifying assumptions of this RDD, and finally presents some results.

A.4.1 Literature

The idea to use an RDD in an electoral context has first been explored by Lee et al. (2004) and Pettersson-Lidbom (2008). While the former analyzes the effect of electoral strength on subsequent roll-call voting, the latter investigates the effect of party ideology on policy-making. The idea has for example also been used by Lee (2008), Ferreira and Gyourko (2009), and Pettersson-Lidbom (2012). Lee (2008) investigates the incumbency advantage in politics, Ferreira and Gyourko (2009) the effect of party control on policies in US cities, and Pettersson-Lidbom (2012) the effect of legislature size on government size.

For nice reviews of papers using close elections RDD, see Caughey and Sekhon (2011) and Snyder et al. (2012). Caughey and Sekhon (2011) argue that elections RDD may be problematic since close winners may often differ in pretreatment covariates compared to close losers in US House elections due to manipulation around the threshold. But they admit that the problem is less severe at the state level where races are often less professionalized. They also propose to check effects on lagged response variables in the RDD (which I do). Snyder et al. (2012), on the other hand, show that covariate imbalances across the election threshold occur even without any sorting around the threshold simply due to the underlying distribution of partisanship in the electorate. According to their view, these imbalances do not pose any problems to elections RDD as long as a polynomial of the forcing variable is included (which I do as well). Furthermore, Eggers et al. (2013) show that problematic imbalances seem to be a US House anomaly during the period after WWII. Investigating more than 40,000 close races in several countries, they do not find imbalances in any other electoral context (including, for example, US statewide and state legislative elections). For general practical RDD introductions, see Imbens and Lemieux (2008) and Lee and Lemieux (2010).

The RDD in this paper is special since it is characterized by several interdependent treatment assignment variables. For theoretical treatments of this and similar topics, see Imbens and Zajonc (2011) and Papay et al. (2011). See Dell (2010) for a recent application to a case with two independent assignment variables.

A.4.2 Multiple Interdependent Assignment Variables

The setting is non-standard since it is characterized by three interdependent assignment variables: the election result of the gubernatorial race and the two seat shares for the two legislative chambers resulting from the legislative elections. These three election results jointly determine if government is divided or unified in a state. When multiple variables are responsible for treatment assignment and only the average treatment effect is of interest, the most straightforward approach is to collapse the multiple variables into one single (artificial) assignment variable taking the value of the one of the original assignment variables which has the value that is closest to the treatment boundary (as suggested by Imbens and Zajonc (2011)). This makes the RDD one-dimensional again by treating the closest distance to the treatment boundary as assignment variable. The new assignment variable therefore measures the closeness of the closest election that could have changed treatment (from divided government to unified or vice versa) if the election had resulted in the other party winning. Since in this analysis the interest indeed lies in identifying the average effect (of divided versus unified government), the assignment variables are collapsed in the described way. The collapsing procedure is explained next.

The goal is to reduce the dimensionality of assignment from three to one to be able to analyze the setting using the standard univariate regression discontinuity framework. Therefore, the three assignment variables have to be collapsed into one. The treatment of interest is divided versus unified government. Unified government refers to a situation when the governor, the majority of legislators in the house, and the majority of legislators in the senate are all from the same party. Divided government refers to all other cases. The three institutions determining treatment therefore are governor, house, and senate. The three variables determining treatment assignment are the election results for these three institutions. We seek to identify exogenous variation in the treatment, i.e. we want to focus the analysis on elections that fulfill two criteria: First, they had the potential to change treatment from divided to unified or vice versa. Second, they were close in the sense that it was not entirely clear to voters beforehand which party would win the election. An election that does *not* fulfill the first criterion would be the election of a state senate when the governor's office and the house are not up for election, the governor is a Republican, and the senate is Democrat (meaning having a Democratic majority). In that case, regardless of the outcome of the house election the future government will be divided since incumbent governor and senate majority are not from the same party. There is no way the house election can produce a unified government. In contrast, all elections where the election has the potential to theoretically result in both treatments fulfill the first criterion, have the potential to provide us with quasi-random treatment assignment, and are included in the regression discontinuity analysis. For all those elections, the closeness to the treatment boundary (where divided changes to unified government or vice versa) is determined and assigned as value of the new assignment variable. The creation of the new assignment variable is illustrated using some examples in the following paragraph.

The most common electoral structure in US states is to elect governor, senate, and house on the same day every four years and to additionally elect the house and the senate (but not the governor) after two years since state legislators are usually elected for two years only. The second type of elections (when only house and senate are elected) are usually called "midterm elections" since they take place in the mid of the term of the governor (who is in office for four years). There are states with different electoral structures and all these different structures are taken into account when coding the new assignment variable, but the just presented structure is by far most common in US states. In the sample from 1978 to 2010, there are more than 400 elections of the first type (governor, house, and senate up for election) and more than 300 elections of the midterm type (house and senate up for election). But there are only between 1 and 40 elections of any other type. The creation of the new assignment variable proceeds in 6 steps for every election day in every state between 1978 and 2010. The steps are:

(1) Check which of the three institutions (governor, house, senate) are up for election on the election day under consideration.
(2) Determine party control of those institutions that are not up for election.
(3) Determine if the election day can potentially change treatment from divided government to unified government or vice versa. If yes, determine which of the elections (of which institutions) can change treatment.
(4) For those elections that can change treatment determine the value of the (multiple) assignment variables, i.e. the election results. For governors, this is the vote margin. For legislatures, this is the deviation of the Democratic seat share from 0.5.[1]
(5) Assign the smallest value of these assignment variables from step (4) that would have been sufficient for a treatment change to the new (to be created) assignment variable.
(6) Extend the new assignment variable as a measure of closeness to the treatment boundary to all following years until the next election takes place.

Let us have a look at some examples for midterm elections. The logic for elections where other combinations of institutions are up for elections is similar.

Example 1: Suppose we are confronted with a standard midterm election day where senate and house are up for election. Let us suppose that the incumbent governor who is not up for election is a Republican. Let us further suppose that both house and senate happen to get a Republican majority in the current election. Government is unified. Clearly, both elections (the house and the senate election) had the theoretical potential of having assigned a divided government treatment instead of a unified government treatment (if they had resulted in the Democrats winning a majority). If in only one of the two legislative elections the Democrats had gained a majority, government would have been divided. The new treatment assignment variable will be assigned the assignment variable of house or senate depending on which election was closer.

Example 2: Suppose we are confronted with a standard midterm election day where senate and house are up for election. Let us suppose that the incumbent governor who is not up for election is a Republican. Let us further suppose that the senate happens to get a Democratic majority in the current election, the house happens to get a Republican majority. Government therefore is divided. In this case, only the senate election had the theoretical potential of changing the treatment to unified government. If the house election had resulted in a Democratic majority, this would have not changed treatment. The new treatment assignment variable will therefore be assigned the assignment variable of the senate.

Example 3: Suppose we are confronted with a standard midterm election day where senate and house are up for election. Let us suppose that the incumbent governor who is not up for election is a Republican. Let us further suppose that both house and senate happen to get a Democratic majority in the current election. Government is divided. Clearly, only both elections together had the potential of having resulted in unified government instead had they *both* resulted in a Republican majority. The new treatment assignment variable will therefore be assigned the sum of the assignment variables of the house and the senate election.

A.4.3 Identifying Assumptions

The important identifying assumptions of the approach are the following. First, there has to be some randomness in final election results. This seems obvious. Second, there must not be

[1]The usually very small number of independent legislators are split equally between Republicans and Democrats when calculating seat shares.

any sorting around the discontinuity, i.e. there must not be any manipulation of election results by candidates close to the threshold. This second assumption is checked by investigating the smoothness of the density of observations around the threshold and by testing the similarity of relevant pretreatment observables across the threshold. Figure A.3 shows that the density of the closeness assignment variable is indeed smooth at zero and manipulation around the cutoff should not be of any concern in this analysis. Appendix Table A.7 and Appendix Table A.8 show that treatment and control group are similar in terms of predetermined covariates when focusing on observations close to the threshold (i.e. close elections).

Figure A.3: Distribution of the RDD Assignment Variable

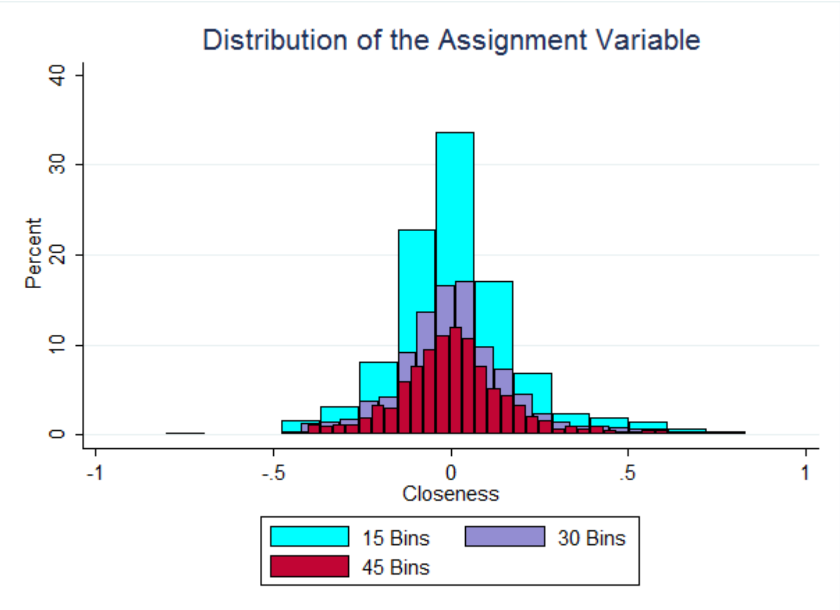


Table A.7: Means by Divided versus Unified Government (Full Sample)

Variable	**Unif. Govt. Mean**	**Div. Govt. Mean**	**ttest Δ p value**
Reform Dummy	0.2014	0.2503	0.0254
Broad Reform Dummy	0.2302	0.2811	0.0256
Broad Reform Count Variable	0.5079	0.6457	0.0784
Narrow Reform Dummy	0.1568	0.1951	0.0546
Narrow Reform Count Variable	0.2187	0.2696	0.1181
Reform Package (including contractive and expansive policies)	0.0669	0.0940	0.2014
Reform Dummy Geographic Neighbors	0.2181	0.2196	0.9136
Reform Dummy Population Size Neighbors	0.2219	0.2263	0.7604
Share of AFDC/TANF Recipients in Pop. (Caseload)	0.0290	0.0306	0.0797
% Unemployed (/1000)	0.0060	0.0060	0.4856
Deflated Total State Revenue per Capita (/1000)	2.0788	2.2430	0.0000
Unmarried Birth (per 1,000 unmarried women)	28.599	28.984	0.4409
Maximum AFDC/TANF Benefit Level for a Family of 3 (/1000)	0.3465	0.3755	0.0000
Per Capita Income (/1000)	22.143	23.067	0.0872
Population (/1000000)	5.2472	5.9231	0.0279
% Population Black	10.878	9.9415	0.0581
% Population Latino	6.4043	7.0629	0.1550
% Population 65 or older	12.421	12.406	0.8759
% Population 17 or younger	26.303	26.021	0.0407
% Immigrant Population	1.6788	2.0732	0.0001
90th/10th Ratio of Household Income	7.9548	8.0187	0.3739
Governor Lame Duck (i.e. cannot be reelected)	0.2950	0.2362	0.0106
Gubernatorial Election	0.2820	0.2734	0.7134
Polarization Senate	0.3077	0.3761	0.0000
Polarization House	0.3190	0.3777	0.0000
% Women in State Legislature	17.433	18.933	0.0006
% Democratic Votes in Last Presidential Election (/1000)	0.0442	0.0448	0.1750
Citizens Ideology (Berry et al. 1998) (/1000)	0.0474	0.0500	0.0014
Democratic Seat Share in Senate	0.5843	0.5436	0.0000
Democratic Seat Share in House	0.5751	0.5478	0.0028
Government Ideology (Berry et al. 1998) (/1000)	0.0493	0.0502	0.5223
Governor Party Dummy (1 = Democrat)	0.6273	0.4159	0.0000
N	695	779	1474

Notes: The first and second column give variable means for the group of unified and the group of divided governments respectively. The third column gives p values from a two sided group mean comparison t test. For details on coding, variables meanings, and data sources, see the Data Appendix.

Table A.8: Means by Divided versus Unified Government (5% Closeness Sample)

Variable	**Unif. Govt. Mean**	**Div. Govt. Mean**	**ttest Δ p value**
Reform Dummy	0.2066	0.2863	0.0487
Broad Reform Dummy	0.2300	0.3105	0.0535
Broad Reform Count Variable	0.4883	0.6613	0.1605
Narrow Reform Dummy	0.1549	0.1976	0.2334
Narrow Reform Count Variable	0.1831	0.2823	0.0819
Reform Package (including contractive and expansive policies)	0.0800	0.0882	0.8341
Reform Dummy Geographic Neighbors	0.2417	0.2337	0.7624
Reform Dummy Population Size Neighbors	0.2402	0.2278	0.6314
Share of AFDC/TANF Recipients in Pop. (Caseload)	0.0298	0.0324	0.1006
% Unemployed (/1000)	0.0058	0.0060	0.2620
Deflated Total State Revenue per Capita (/1000)	2.0997	2.1353	0.5301
Unmarried Birth (per 1,000 unmarried women)	28.765	27.455	0.1150
Maximum AFDC/TANF Benefit Level for a Family of 3 (/1000)	0.3611	0.3801	0.1436
Per Capita Income (/1000)	22.251	21.974	0.7551
Population (/1000000)	5.5576	6.3102	0.1492
% Population Black	10.038	9.0870	0.2600
% Population Latino	6.4436	6.4143	0.9725
% Population 65 or older	12.380	12.565	0.2316
% Population 17 or younger	26.115	26.089	0.9075
% Immigrant Population	1.6561	1.8487	0.2867
90th/10th Ratio of Household Income	7.8632	7.8703	0.9516
Governor Lame Duck (i.e. cannot be reelected)	0.2300	0.1573	0.0475
Gubernatorial Election	0.1831	0.1734	0.7863
Polarization Senate	0.3749	0.3891	0.1562
Polarization House	0.3914	0.3894	0.8324
% Women in State Legislature	18.948	18.077	0.2654
% Democratic Votes in Last Presidential Election (/1000)	0.0441	0.0448	0.2837
Citizens Ideology (Berry et al. 1998) (/1000)	0.0485	0.0507	0.1027
Democratic Seat Share in Senate	0.5412	0.5587	0.2145
Democratic Seat Share in House	0.5396	0.5592	0.1333
Government Ideology (Berry et al. 1998) (/1000)	0.0504	0.0500	0.8597
Governor Party Dummy (1 = Democrat)	0.5915	0.3831	0.0000
N	213	248	461

Notes: The sample is restricted to observations where the election result determining whether government would be divided or unified was decided by a 5 percentage points or smaller vote/seat margin. For details, see the RDD Appendix. The first and second column give variable means for the group of unified and the group of divided governments respectively. The third column gives p values from a two sided group mean comparison t test. For details on coding, variables meanings, and data sources, see the Data Appendix.

A.4.4 Results

This subsection presents the results from a short RDD analysis. The low number of observations makes a parametric approach preferable over a nonparametric approach. The regressions therefore fit a polynomial in the collapsed assignment variable to estimate the treatment effect at the boundary. Restricting the analysis to close elections (which is a nonparametric technique) makes the analysis semiparametric. In that sense the approach could be best described as a semiparametric RDD with multiple interdependent assignment variables.

Appendix Table A.9 shows the results. Columns (1) to (4) have the welfare reform dummy as dependent variable. Columns (5) to (8) have the lagged reform dummy as dependent variable, i.e. the right part of the table presents a placebo test where no treatment effect of divided government is expected (since the current form of government should not affect reform adoption in the past). Each triple of numbers (coefficient, standard error, R squared) shows the result of one RDD regression of reform on divided government. Columns (1) to (4) and (5) to (8) add a polynomial of degree 0 to 3 in the assignment variable as control function. The first two rows look at the full sample, the third and fourth restrict the sample to observations where the collapsed treatment assignment variables takes values of 5% or smaller, i.e. to close elections (where the distance to the treatment boundary is 5 percentage points or less). Restricting the sample mimics a nonparametric approach. A fully nonparametric approach is infeasible because of the small number of observations. While the first and the third row regressions do not include any controls besides the polynomial control function, the second and the fourth row regressions include state and year fixed effects, state specific linear trends, and the full set of demographic conditions (lagged by one year) as controls. If the RDD is indeed quasi-random and the identifying assumptions are fulfilled, controls are not necessary for identification and should not change much in terms of results. However, it has been argued that including them can increase the precision of the estimates and may be especially worthwhile in the case of a low number of observations (Hoxby (2000), Pettersson-Lidbom (2008), and Pettersson-Lidbom (2012)).

As can be seen from the table, the RDD in columns (1) to (4) largely confirms the results from the panel data regressions presented before. There is a significantly positive effect of divided government on the adoption of reforms. Although the size of the RDD coefficients is not directly comparable to the results from the analysis before since the RDD estimates the effect at the treatment boundary, the analysis nevertheless clearly supports the finding that divided governments are more likely to reform than unified governments. Adding controls to the RDD seems to increase precision, but as expected does not change the overall picture. The placebo check in columns (5) to (8) also works out: It seems reasonable to assume that the current form of government affects current and possibly future reform adoption, but cannot affect reform adoption in the past. If the empirical design is valid, one should therefore not find any effect of divided government on previous year reform. This is indeed that the right part of Appendix Table A.9 shows: The esimated coefficients are rather small and none of them is significant at any conventional level.

Table A.9: Divided Government and Reform (RDD Analysis)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
	Div. Govt. and Reform				Div. Govt. and Previous Year Reform (Placebo)				
Polynomial	**0**	**1**	**2**	**3**	**0**	**1**	**2**	**3**	**Polynomial**
FULL SAMPLE									**FULL SAMPLE**
No controls	0.0489*	0.0528*	0.0465	0.0476*	0.0171	0.0200	0.0172	0.0187	**No controls**
1,449 observations	(0.0287)	(0.0283)	(0.0283)	(0.0282)	(0.0290)	(0.0289)	(0.0296)	(0.0295)	1,453 observations
R-squared	0.003	0.004	0.008	0.010	0.000	0.001	0.002	0.005	R-squared
With controls	0.0649***	0.0575**	0.0562**	0.0567**	0.0133	0.0108	0.0113	0.0106	**With controls**
1,320 observations	(0.0229)	(0.0236)	(0.0235)	(0.0235)	(0.0206)	(0.0215)	(0.0217)	(0.0218)	1,365 observations
R-squared	0.326	0.327	0.329	0.329	0.321	0.322	0.323	0.323	R-squared
5% SAMPLE									**5% SAMPLE**
No controls	0.0797*	0.0809*	0.0808*	0.0826*	-0.0166	-0.0144	-0.0134	-0.00930	**No controls**
461 observations	(0.0447)	(0.0453)	(0.0452)	(0.0463)	(0.0476)	(0.0480)	(0.0480)	(0.0484)	469 observations
R-squared	0.008	0.011	0.011	0.011	0.000	0.010	0.010	0.011	R-squared
With controls	0.106**	0.106**	0.107**	0.108**	0.0261	0.0257	0.0233	0.0234	**With controls**
424 observations	(0.0497)	(0.0501)	(0.0492)	(0.0486)	(0.0755)	(0.0736)	(0.0748)	(0.0751)	437 observations
R-squared	0.463	0.463	0.463	0.464	0.466	0.467	0.469	0.469	R-squared

Notes: The left part of this table shows the effect of divided government on reform. The right part of this table shows the effect of divided government on previous year's reform (placebo test where no effect is expected). Each triple of numbers (coefficient, standard error, R-squared) refers to one estimation. The method of analysis is parametric RDD (with polynomials in the assignment variable from orders 0 to 3) conducted either using the full sample or a sample restricted to cases where the closeness of a hypothetical treatment assignment change from divided to unified government or vice versa was ex post 5 percentage points or smaller (i.e. where the assignment variable takes values of 5% or smaller). Thus, the 5% sample only includes observations where ex post a difference of 5 percentage points or less in the gubernatorial vote margin or in legislature seat shares would have been sufficient for a treatment change from divided to unified government or vice versa. Controls are year and state fixed effects, state specific linear trends, and demographic variables lagged one year. The dependent variable in specification (1) to (4) is a reform dummy that is equal to one if one or more welfare reforms have been introduced in a given state and year. The dependent variable in specifications (5) to (8) is the same variable lagged. Divided Government is a dummy that is equal to one when either the majority of the state's lower legislative chamber or the majority of the state's upper legislative chamber is from another party than the governor. For details regarding any of the variables, see the Data Appendix. Robust standard errors clustered at the state level are shown in parentheses. *** $p<0.01$, ** $p<0.05$ and *$p<0.1$.

A.5 Data

Note that most of the data come directly from Bernecker and Gathmann (2013). Summary statistics are provided in Table 2.1, means of variables by divided versus unified government in Appendix Table A.7.

A.5.1 Divided Government Variables

Divided Government is a dummy that is equal to one when either the majority of the state's lower legislative chamber or the majority of the state's upper legislative chamber is from another party than the governor. "Divided Government at Federal Level" is defined equivalently for the US federal government. Split Branch is equal to one if the governor in a state is confronted with majorities of the opposing party in both legislative chambers. Split Legislature is equal to one if the majorities in the two legislative chambers in a state are from opposing parties. All divided government measures exclude cases where the governor is party-independent (neither a Democrat nor a Republican) and cases where a legislative chamber is itself split, i.e. having the same amount of Democratic and Republican seats. The variables "Divided Government with past..." are defined on the level of governments (instead of years) and condition on the past type of government, e.g. "Divided Government with past Republican Unified Government" refers to a divided government that replaces a Republican unified government which has been in office before. The two variables "New Divided Government via General Elections" and "New Divided

Government via Midterm Elections" are also defined on the level of governments. The former is a dummy that is equal to one if there was a change from unified to divided government and the governor was up for election in the most recent election. The latter is a dummy that is equal to one if there was a change from unified to divided government and the governor was not up for election in the most recent election. Note that the change from a unified Democratic to a divided government typically occurs via general elections (often via a Republican governor being elected). The change from a unified Republican to a divided government, on the other hand, usually occurs via midterm elections (via the Republicans losing the majority in one of the legislative chambers). These patterns result in highly significant and sizeable correlations between "Divided Government with past Democratic Unified Government" and "New Divided Government via General Elections" (corr=0.75) on the one hand and "Divided Government with past Republican Unified Government" and "New Divided Government via Midterm Elections" (corr=0.4) on the other hand. This is also the reason why means and standard deviations of these dummy variables are very similar. A careful inspection reveals, however, that the variables are in fact not only theoretically, but also empirically not the same. The underlying data for the divided government variables have been obtained from Klarner (2003) and Klarner's webpage (http://www.indstate.edu/polsci/klarnerpolitics.htm).

A.5.2 Welfare Reform Variables

The reform dummy used as dependent variable in most of the analyses in this paper is constructed from different data sources. From 1978 to the Welfare Reform in 1996, the reform dummy is equal to one if a state has filed a waiver application at the federal level in a given state and year. This includes waiver applications rejected by the federal level and also applications referring to only some counties of the state. The data on waivers have been obtained from Lieberman and Shaw (2000) and cross-checked using Koerper (1996) and Crouse (1999). The waiver data do not go back to the period before 1978. However, states have in principle been in a position to file waiver applications already since 1962. But since waivers did not become popular before the late 1970s, the data starting in 1978 nevertheless capture most relevant welfare waiver activity in the states.

With the implementation of the Welfare Reform in 1996, welfare waivers became irrelevant. Within some federal guidelines, states were now free to set their own welfare policy rules. After 1996, the reform dummy is equal to one if a state changed its welfare policy rules. Information on welfare policy changes at the state level is obtained from the Welfare Rules Database maintained by the Urban Institute (Urban Institute (2012)). This database contains the states' welfare rules from 1996 onwards. The basis of the data are the plans of the states (approved by governor and legislature). Data collection is done via states' caseworker manuals and policy updates that are sent to them during the year. TANF administrators of the states verify the data on the policy rules before publication. This makes the data most complete and up-to-date and "one of the most reliable sources on TANF social policies available" (Fellowes and Rowe (2004), p. 365).

The baseline reform dummy is based on changes in 14 highly relevant policy rules in the areas family caps, work requirements, time limits, and sanctions. The detailed rules are given in Appendix Table A.10. The broad reform dummy is based on these 14 and 10 additional policy rules (24 rules in total). These additional rules cover eligibility tests and requirements and are presented in Appendix Table A.11. See Bernecker and Gathmann (2013) for further coding details regarding the policy rules. The narrow reform dummy captures only contractive reforms, is based on only a subset of the baseline rules, and in some cases also uses stricter codings (to cover significant reforms only; see modification in parentheses): family caps, hours requirement (of at least 30 hours), work upon enrollment, reapply, duration of lifetime limit (below 60 months), benefit reduction after intermittent time limit (removal of whole family benefit). The

Table A.10: Welfare Policy Rules used for Definition of Reforms (1996-2010)

Policy Rule	Description of Rule
Family Cap (1)	Benefits do not increase if an additional child is born in family while receiving benefits.
Work Requirements (4)	
Hours Requirement	Minimum # of hours a recipient must participate in work-related activities.
Work upon Enrollment	Work requ. apply at application stage, approval stage or upon benefit receipt (or later).
Time Limit to Work	Work at least 20 hours per week in unsubsidized job after certain period of benefit receipt.
Exemptions to Work Requirements	Number of exemptions state allows to Hours Requ. (for older, ill beneficiaries etc.).
Time Limits (4)	
Duration of Lifetime Limit	Maximum # of months an assistance unit can receive benefits over the lifetime.
Intermittent Time Limit	# months time an assistance unit can receive benefits without interruption.
Benefit Cut after Interm. Time Limit	How much benefits are reduced when assistance unit hits the intermittent time limit.
Time Limit Extensions	Whether the state offers any type of time limit extension.
Sanctions (5)	
Severity Worst Sanction	How much benefits are reduced under worst sanction for non-compliance with work requ.
Duration Worst Sanction	Length of most severe sanction for not complying with work requirements.
Reapply	Whether the unit has to reapply after worst sanction for non-compliance has been imposed.
Severity Initial Sanction (year≥1999)	How much benefits are reduced for first non-compliance with work requirements.
Duration Initial Sanction (year≥1998)	Length of first sanction for not complying with work requirements.

Notes: See the Data Appendix on further details. Source: Own calculations based on Urban Institute's Welfare Rules Database (Urban Institute 2012).

narrow reform dummy also includes "mandatory job search prior to TANF eligibility" and "worst sanction is removal of full family benefit" as two additional policy rules (8 rules in total). The reform package dummy, finally, is only defined for 1996 to 2010 and is equal to one when at least one contractive and at least one expansive policy rule change occur in the same year. The reform package dummy is based on the baseline set of 14 rules (but uses a dummy whether the first sanction is the removal of the full family benefit instead of the two rules referring to severity and duration of the initial sanction since the latter two are only available from 1999 onwards).

For the analysis of spillover effects, the reform dummy in neighboring states is relevant. The Reform Dummy Geographic Neighbors is the average of the Reform Dummy for all geographically adjacent states. For each state, it thus measures the share of neighboring states that have conducted a reform in a given year. The Reform Dummy Population Size Neighbors does the same, but considers states with a similar population size instead of geographic neighbors. For this second measure, the states are divided into ten different bands according to their population size in 1978. The ten bands are (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), (ND DE VT WY AK).

Data on the Maximum AFDC/TANF Benefit Level for a Family of Three (with no income) which is used as a control in parts of the analysis has been obtained from Han et al. (2009) who made their data available at http://www.nber.org/workfamily/ and updated using data provided in the Welfare Rules Database maintained by the Urban Institute (Urban Institute (2012)) available at http://anfdata.urban.org/wrd/tables.cfm (Table II.A.4).

A.5.3 Demographic Variables

The Share of AFDC/TANF Recipients in Population (Caseload) is taken from Moffitt (2002a) until 1998 and updated to 2010 using the Statistical Abstract (United States Census Bureau (2011)). The % Unemployed is also taken from Moffitt (2002a) until 1998 and updated to 2010 using the website of the Bureau of Labor Statistics. Per Capita Income is taken from the

Table A.11: Additional Welfare Policy Rules Used for Broad Definition of Reforms (1996-2010)

Policy Rule	Description of Rule
Asset Test for Eligibility (2)	
Monetary Assets Allowed	Value of maximum unrestricted assets family may have and still be eligible for cash benefits.
Vehicle Exemption	Monetized value of vehicle exemptions that are not counted towards the assets test for applicants.
Eligibility Requirements (2)	
Eligibility of Pregnant Women	Whether a state allows pregnant women with no other children to be eligible.
Eligibility of Minor Parent	Whether a minor parent (under age 18 and never married) is eligible to head a TANF assistance unit.
Two-Parent Families Eligibility (2)	
Two-Parent Work Hrs Limit	Limit on working hours a state imposes on two-parent households to stay eligible.
Two-Parent Wait Period	Whether state requires principal wage earner to have worked for specified time period for eligibility.
Additional Requirements (4)	
School Requirement	Whether state mandates children to attend school or maintain a certain GPA.
School Bonus	Whether a state offers financial incentives to assistance units when children meet state set standards.
Immunization Requirement	Whether state mandates children in the assistance unit to meet standard immunizations for children.
Other Health Requirement	Whether state requires adults and children to get regular health checkups.

Notes: See the Data Appendix on further details. Source: Own calculations based on Urban Institute's Welfare Rules Database (Urban Institute 2012).

website of the Bureau of Economic Analysis and deflated by the urban consumer price index (with year 2002=100). The variables Population, % Black Population, % Population 65 or older, % Population 17 or younger are all taken from the Statistical Abstract (United States Census Bureau (2011)). The % Population Latino has been obtained from several websites of the US Census Bureau: http://www.census.gov/popest/data/historical/1980s/state.html (for the 1980s), http://www.census.gov/popest/data/state/asrh/1990s/st_race_hisp.html (for the 1990s), and http://www.census.gov/popest/data/intercensal/state/state2010.html (for the 2000s). The % Immigrant Population refers to legal immigrants admitted by state of intended residence (then divided by state population) 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)) for 2011 and for previous years (available at http://www.dhs.gov/yearbook-immigration-statistics). Unmarried Birth refers to the % of all 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. For the remaining years, data have been obtained from the Centers for Disease Control and Prevention and the National Vital Statistics System (available at http://www.cdc.gov/nchs/data_access/vitalstats/VitalStats_Births.htm and http://205.207.175.93/VitalStats/) and completed and cross-checked using data available at the National Bureau of Economic Research: http://www.nber.org/data/vital-statistics-natality-data.html. The 90th/10th Ratio of Household Income (90th percentile divided by 10th percentile of all positive household incomes) is calculated from the March Current Population Survey (Center for Economic and Policy Research (2012)).

A.5.4 Political Variables

The % Democratic Votes in Last Presidential Election is taken from the Statistical Abstract (United States Census Bureau (2011)) and updated using Leip (2012). The Democratic Seat Share in Upper House and the Democratic Seat Share in Lower House are calculated based on information about the number of legislators by party and the total number of seats of state legislatures obtained from Klarner (2003) and Klarner's webpage (http://www.indstate.edu/polsci/klarnerpolitics.htm). This is also the source for the Governor Party variable. Polarization of Senate and House are calculated as 0.5-|democratic seat share-0.5|. The % Women in State Legislature is obtained from the website of the Center for American Women and Politics (Center for American Women and Politics (2012)). Governor Lame Duck is equal to one if the incumbent governor cannot run for reelection. Gubernatorial Election is a dummy equal to one if a gubernatorial election took place this year. Both variables are obtained from List and Sturm (2006) until 2000 and updated using Leip (2012). Citizens Ideology and Government Ideology are calculated by Berry et al. (1998) from 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. assign an ideology rating to the citizens of each congressional district using a weighted average of the score of the congressional member and his or her election opponent, weighting the scores according to the number of votes they 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 government ideology is constructed by assigning to the governor and major party delegations in the legislature the ratings of the members of Congress from their party. Updates of these ideology data are available at http://www.bama.ua.edu/ rcfording/stateideology.html.

A.5.5 Public Finance Variable

The variable Deflated Total State Revenue Per Capita is calculated by using data on state revenues (obtained from Paul Ehmann at the US Census Bureau (http://www.census.gov/govs/state/)) and dividing those numbers by the state population (see demographic variables explained above) and by the urban consumer price index (with years 1982-1984=100) provided by the US Bureau of Labor Statistics at ftp://ftp.bls.gov/pub/special.requests/cpi/cpiai.txt.

Appendix B

Appendix to Chapter 3

B.1 Data Appendix

B.1.1 Welfare Waivers and Policy Rules

To obtain measures of policy experimentation and reversals for the post-1996 period, we rely on the Welfare Rules Database by the Urban Institute. The Welfare Rules Database (WRD) was developed to provide detailed information about states' TANF policies obtained from caseworker manuals and regulations. The information here is typically more detailed and up-to-date than the official plans states submit periodically to the federal government (providing an overview of states' choices under the block grant). Our main analysis focuses on a set of rules in policy areas that were at the center of the public and political debate surrounding welfare reform. See Table B.1 for an overview.

Table B.1: Policy Rules Used to Define Experimentation and Reversals (1996-2010)

Policy Rule	Description of Rule
Family Cap (1)	Benefits do not increase if an additional child is born in family while receiving benefits.
Work Requirements (4)	
Hours Requirement	Minimum # of hours a recipient must participate in work-related activities.
Work upon Enrollment	Work requ. apply at application stage, approval stage or upon benefit receipt (or later).
Time Limit to Work	Work at least 20 hours per week in unsubsidized job after certain period of benefit receipt.
Exemptions to Work Requirements	Number of exemptions state allows to Hours Requ. (for older, ill beneficiaries etc.).
Sanctions (4)	
Severity Initial Sanction	Whether initial sanction for non-compliance is removal of full family benefit.
Severity Worst Sanction	How much benefits are reduced under worst sanction for non-compliance with work requ.
Duration Worst Sanction	Length of most severe sanction for not complying with work requ.
Reapply	Whether unit has to reapply after worst sanction for non-compliance has been imposed.
Time Limits (4)	
Intermittent Time Limit	# months time an assistance units can receive benefits without interruption.
Benefit Cut after Interm. Time Limit	How much benefits are reduced when assistance units hits interm. time limit.
Duration of Lifetime Limit	Maximum # of months an assistance unit can receive benefits over the lifetime.
Time Limit Extensions	Whether the state offers any type of time limit extension.

Notes: See data appendix on further details how the policy rules were coded. Sources: Authors' calculations based on Urban Institute's Welfare Rules Database (Urban Institute, 2012).

Family Caps. Under AFDC, benefit levels increased with family size. Hence, when a child was born to a member of an assistance unit, the benefit increased to meet the needs of the new child. Under a family cap policy, the benefit increase that an assistance unit would normally receive for adding another member to the unit will be limited. 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.

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 after a given period. Work programs vary widely from state to state including 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 fulltime employment. The second rule defines whether the work requirement applies after several months of benefit receipt or by the time of application/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.

Sanctions. The Job Opportunities and Basic Skills Training Program (JOBS), part of the 1988 Family Support Act, provides education, training and work experience activities to AFDC recipients. Under TANF, states now require most adult heading in an assistance unit to perform some type of work-related activity. If a benefit unit does not comply with these requirements, states can impose 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 we code 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. The federal government has itself imposed a 60 months lifetime limit; states that wish to extend benefit receipt beyond five years have to use state funds for it. The final rule defines whether the state allows for any extensions to the lifetime limit or not.

B.1.2 Politics and Ideology Measures

Governor Characteristics. Information on US governors comes from the website of the National Governor's Association and merged with data kindly provided by David J. Andersen from the Eagleton Institute of Politics at Rutgers University. The quality of a governor is calculated as years in political office, more precisely as the number of years between taking the first electoral office (such as member of the State Senate or Attorney General) and the inauguration as governor. This gives a continuous quality measure. The low quality dummy is equal to one for all governors with this quality measure being below the median (and zero otherwise). Information on gubernatorial elections and term limits are taken from List and Sturm (2006); Council of State Governments (2012); and Leip (2012)).

State Legislature. The data for the party composition of the state legislature, the party of the governor and indicators for a divided government is obtained from Klarner (2003) and updated using Klarner's webpage (http://www.indstate.edu/polsci/klarnerpolitics.htm). Polarization 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. Information on women legislators and women governors was collected from the website of the Center for American Women and Politics (Center for American Women and Politics (2012)). Data for women legislators are available biannually from 1975 to 1983 and annually thereafter. The data for African American legislators for 1984 to 1993 is from Preuhs (2006) and for 1994 to 2005 from the website of the National Conference on State Legislatures. Data on term limits and vote margins for governors from 1960 to 2000 are taken from List and Sturm (2006) and updated using the Book of the States and Leip (2012). Data for term limits of state legislators are from the National Conference on State Legislatures.

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). We also use voter and government ideology 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 using a weighted average of the score of the congressional member and his or her election opponent, weighting the scores according to the number of votes they 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 government ideology is constructed by assigning to the governor and major party delegations in the legislature the ratings of the members of Congress from their party. Updates of these ideology data are available at http://www.bama.ua.edu/∼rcfording/stateideology.html.

B.1.3 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)). We collect data on the Latino population from several websites of the US Census Bureau:
http://www.census.gov/popest/data/historical/1980s/state.html (1980-89),
http://www.census.gov/popest/data/state/asrh/1990s/st_race_hisp.html (1990-1999), and http://www.census.gov/popest/data/intercensal/state/state2010.html (2000-2010). The size of the immigrant population refers to the number of legal immigrants admitted by state of intended residence 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)) which is available online at http://www.dhs.gov/yearbook-immigration-statistics. Personal income per capita, transfer receipts, wage income and total employment is taken from the website of the Bureau of Economic Analysis. The unemployment rate for 1960 to 1998 is from Moffitt (2002a) and updated for 1998 to 2010 from the website of the Bureau of Labor Statistics. Indicators of income inequality (standard deviation, 90th/10th percentile ratio and Gini coefficient of total family income, personal income for men and women as well as wage and salary payments for men

and women) are 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 all 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 (available at http://www.cdc.gov/nchs/data_access/vitalstats/VitalStats_Births.htm and http://205.207.175.93/VitalStats/) and Vital Statistics available from the National Bureau of Economic Research (http://www.nber.org/data/vital-statistics-natality-data.html).

State AFDC/TANF Caseloads and Finances. TANF caseloads for both fiscal year and calendar year are taken from Moffitt (2002a) for 1960 to 1998 and updated to 2010 using the Statistical Abstract (United States Census Bureau (2011)). Average monthly payments for TANF recipients and families (by fiscal year) are also from Moffitt (2002a) for 1960-1997 and updated to 2010 using the Annual Statistical Supplement to the Social Security Administration. Data on the Maximum AFDC/TANF benefit level for a family of three (with no other income) comes from Han et al. (2009) and can be downloaded from http://www.nber.org/workfamily/. We update the series to 2010 using the Welfare Rules Database by the Urban Institute (Urban Institute (2012)) available at http://anfdata.urban.org/wrd/tables.cfm (Table II.A.4). Job entry, job retention and earnings gain data for 1997-2010 are taken from the High Performance Bonus data (table 5.4) reported in the TANF-Annual Reports to Congress (U.S. Department of Health and Human Services, Administration for Children and Families (2009)).

To control for the current financial situation in a state, we obtain data on state revenues, debt outstanding and state expenditures (in thousands of US Dollars) from Paul Ehmann at the US Census Bureau (http://www.census.gov/govs/state/). All fiscal variables are then converted into real per capita measures using the state population variable described above and the urban consumer price index (with years 1982-1984=100) from the Bureau of Labor Statistics (ftp://ftp.bls.gov/pub/special.requests/cpi/cpiai.txt).

Spillover Effects across States. For the analysis of spillover effects, welfare experiments in neighboring states are relevant. For each state, the Geographic Neighbors' Experiments variable refers to the average number of waivers for all states geographically adjacent to this state. The Population Size Neighbors' Experiments measure is similar, but considers states with a similar population size instead of geographic neighbors. For this second measure, the states are divided into ten different bands according to their population size in 1978. The ten bands are (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), (ND DE VT WY AK).

Appendix C

Appendix to Chapter 4

C.1 Appendix Tables

Table C.1: Effect of 2005 Vote Margin on 2009 Vote Margin

Dependent variable: Vote margin in the 2009 election				
	(1)	(2)	(3)	(4)
Vote margin in the 2005 election	0.638***	0.669***	0.664***	0.664***
	(0.0509)	(0.0344)	(0.0338)	(0.0336)
Social Democrat		-15.36***	-15.12***	-14.92***
		(0.551)	(0.586)	(0.628)
Demographic Controls	NO	NO	YES	YES
Political Controls	NO	NO	NO	YES
Observations	162	162	162	162
R-squared	0.416	0.880	0.883	0.886

Notes: Sample includes district MPs of Social Democratic and Christian Democratic Party who have been members of both the parliament elected in 2005 and the parliament elected in 2009 (excluding early leavers and late entrants). Dependent variable is the vote margin in the 2009 election. Controls are the same as in Table 1 (demographic: female, age, partner, children, university, phd; political: number of terms, minister, leading position). All regressions include a constant. Robust standard errors are reported in parentheses. *, **, *** denote significance at the 10, 5, 1 % level.

Table C.2: Summary Statistics for District MPs of Social Democratic and Christian Democratic Party in Parliament Elected in 2009

Variable	Mean	Std. Dev.	Description	Source
% Absent days	7.74	9.29	% of days with at least one missed vote	own calculations; Bundestag (2012)
% Absent days w/o excuse	4.49	6.21	% of unexcused days with at least one missed vote	own calculations; Bundestag (2012)
% Absent votes	6.49	8.13	% of votes missed	own calculations; Bundestag (2012)
Vote margin	11.24	8.42	votes distance to district runner-up candidate in percentage points	Bundeswahlleiter (2012)
Second vote share	33.80	4.56	votes to own party in electoral district in percentage points	Bundeswahlleiter (2012)
Social Democrat	.27	.44	1 if being member of Social Democratic Party	Bundestag (2012)
Female	.21	.41	1 if being female	Bundestag (2012)
Age	50.96	9.05	age in year 2010	Bundestag (2012)
Partner	.74	.44	1 if having partner	Bundestag (2012)
Children	.66	.48	1 if having children	Bundestag (2012)
University	.84	.37	1 if having university degree other than phd	Bundestag (2012)
PhD	.21	.41	1 if having phd	Bundestag (2012)
Terms	2.96	1.95	number of legislative periods served in parliament	Bundestag (2012)
Minister	.04	.20	1 if minister in Angela Merkel's 2009-2013 cabinet	Bundestag (2012)
Leading position	.27	.44	1 if having a leading position in parliament*	Bundestag (2012)
Safe	.36	.48	1 if would have been elected via the party list if lost district	own calculations; Bundeswahlleiter (2012)
% Platform Answers	65.90	38.38	% answered questions at abgeordnetenwatch.de	Abgeordnetenwatch.de (2011)

Notes: The number of observations is 228. All three absence measures refer to sessions between 27/10/2009 and 29/06/2013. All variables are defined on the MP level. *The following positions are considered for the leading position dummy: chairman or vice chairman of the parliament, a committee, a party, a faction, faction manager, party general secretary. Summary statistics for list MPs or MPs in the parliament elected in 2005 are available from the author upon request.

Table C.3: Effect of Vote Margin on Absent Days (Including Christian Social Democrats)

Dependent variable: % Absent days	(1)	(2)	(3)	(4)
Vote margin * Social Democrat		0.744***	0.731***	0.788***
		(0.270)	(0.275)	(0.297)
Vote margin	0.0977	0.0192	0.0317	-0.0715
	(0.0773)	(0.0802)	(0.0775)	(0.0726)
Social Democrat	7.549***	1.980	2.069	1.296
	(1.637)	(2.003)	(2.113)	(2.274)
Christian Social Democrat	1.996	3.143	2.985	2.840
	(2.631)	(2.656)	(2.495)	(2.387)
Demographic Controls	NO	NO	YES	YES
Political Controls	NO	NO	NO	YES
Observations	272	272	272	272
R-squared	0.095	0.134	0.206	0.300

Notes: Sample includes district MPs of Social Democratic, Christian Democratic, and Christian Social Party of the parliament elected in 2009 (excluding early leavers and late entrants). Dependent variable is the percentage of days when the MP missed at least one recorded vote (out of all days with a recorded vote). Controls are the same as in Table 1 (demographic: female, age, partner, children, university, phd; political: number of terms, minister, leading position). All regressions include a constant. Robust standard errors are reported in parentheses. *, **, *** denote significance at the 10, 5, 1 % level.

Table C.4: Effect of Vote Margin on Absent Days (First Stage)

Dependent variable:	(1a) Vote margin	(1b) Vote margin*SD	(2a) Vote margin	(2b) Vote margin*SD	(3a) Vote margin	(3b) Vote margin*SD	(4a) Vote margin	(4b) Vote margin*SD
2nd vote share 2009	1.467***	-0.00269						
	(0.0949)	(0.0366)						
2nd vote share 2009*SD	-0.458***	0.981***						
	(0.174)	(0.0669)						
2nd vote share 1990			1.046***	0.00277	1.0178***	0.0144	1.152***	0.0117
			(0.0741)	(0.0270)	(0.0811)	(0.0259)	(0.0863)	(0.0281)
2nd vote share 1990*SD			-0.384***	0.669***	-0.360**	0.636***	-0.322*	0.702***
			(0.124)	(0.0451)	(0.141)	(0.0451)	(0.166)	(0.0542)
Social Democrat (SD)	13.156**	-23.837***	13.007**	-22.180***	11.794*	-20.962***	8.172	-24.359***
	(5.686)	(2.193)	(5.575)	(2.031)	(6.308)	(2.016)	(7.419)	(2.420)
Female	-0.174	0.211	-0.287	0.0845	-0.154	0.240	0.352	0.205
	(0.884)	(0.341)	(0.930)	(0.339)	(1.026)	(0.328)	(1.131)	(0.369)
Age	-0.0215	0.0201	-0.0253	0.00956	-0.0104	0.0157	-0.103*	-0.00374
	(0.0475)	(0.0183)	(0.0514)	(0.0187)	(0.0560)	(0.0179)	(0.0593)	(0.0193)
Partner	1.936**	0.971**	1.226	0.306	1.036	0.196	-0.772	-0.0714
	(0.969)	(0.374)	(1.013)	(0.369)	(1.010)	(0.351)	(1.165)	(0.380)
Children	-1.829**	-0.177	-0.895	0.428	-0.834	0.556	-0.600	0.497
	(0.917)	(0.356)	(0.957)	(0.349)	(1.060)	(0.339)	(1.070)	(0.349)
University	-0.359	0.308	0.112	0.538	0.547	1.0265***	1.916*	0.825**
	(0.972)	(0.375)	(1.053)	(0.385)	(1.116)	(0.357)	(1.146)	(0.374)
PhD	1.080	0.392	-0.184	0.312	0.0510	0.275	-1.969*	0.0748
	(0.902)	(0.348)	(0.950)	(0.346)	(1.074)	(0.343)	(1.108)	(0.362)
Number of terms	0.714***	0.110	0.724***	0.0844	0.956***	-0.0719	1.122***	0.0797
	(0.211)	(0.0815)	(0.229)	(0.0834)	(0.331)	(0.106)	(0.342)	(0.112)
Minister	0.877	-0.475	2.037	-0.311	1.710	-0.191	1.945	-0.203
	(1.857)	(0.716)	(1.923)	(0.701)	(2.350)	(0.751)	(2.507)	(0.818)
Leading position	1.977**	-0.0416	2.180**	0.209	1.772*	0.123	1.879*	0.107
	(0.793)	(0.306)	(0.841)	(0.306)	(0.981)	(0.314)	(0.989)	(0.322)
F statistic	35.27	73.78	30.30	78.85	23.26	74.24	26.11	56.34
Observations	228	228	218	218	193	193	124	124
R-squared	0.663	0.805	0.640	0.822	0.608	0.832	0.738	0.860

Notes: Each 2SLS regression in Table 2 has two corresponding first stage regressions in this table. Dependent variables are the vote margin in the 2009 election and the vote margin in the 2009 election interacted with a Social Democrat dummy. Samples for the columns are the same ones as the ones for the respective 2SLS regressions in Table 2. All regressions include a constant. Robust standard errors are reported in parentheses. *, **, *** denote significance at the 10, 5, 1 % level.

Appendix D

Appendix to Chapter 5

D.1 Strength of Anchoring in Surveys 1 and 2

Table D.1: Anchoring Can Cause up to 50% of Status Quo Bias (Survey 1)

Band around anchor/status quo	Respondents fraction within band: No Info	Anchoring	Status Quo	Chi square p-value	Relative strength anchoring effect
1%	0.000	0.231	0.485	0.000	0.476
5%	0.030	0.246	0.515	0.000	0.445
10%	0.030	0.262	0.545	0.000	0.449
20%	0.030	0.292	0.576	0.000	0.480
30%	0.152	0.385	0.636	0.000	0.481
50%	0.152	0.508	0.727	0.000	0.619

Notes: Data from Survey 1. 131 observations. Column one gives the +/- range by which a respondent's stated policy bliss point can deviate from the anchor (anchoring group) respectively status quo (no info and status quo group) to still be considered close. For example, 23.1% of respondents in the anchoring group stay within a +/- 1% range of the anchor. Column four gives the p-value of a chi square test checking whether the fraction of respondents within the band is equal across the three groups. The last column gives a rough estimate of the strength of the anchoring effect relative to the status quo bias by calculating: (fraction anchoring - fraction no info) / (fraction status quo - fraction non info).

Table D.2: Anchoring Can Cause up to 50% of Status Quo Bias (Survey 2)

Band around anchor/status quo	Respondents fraction within band: No Info	Anchoring	Status Quo	Chi square p-value	Relative strength anchoring effect
1%	0.017	0.121	0.351	0.000	0.311
5%	0.017	0.138	0.351	0.000	0.363
10%	0.033	0.172	0.368	0.000	0.415
20%	0.033	0.224	0.404	0.000	0.515
30%	0.067	0.310	0.596	0.000	0.460
50%	0.117	0.466	0.667	0.000	0.634

Notes: Data from Survey 2. 171 observations. Column one gives the +/- range by which a respondent's stated policy bliss point can deviate from the anchor (anchoring group) respectively status quo (no info and status quo group) to still be considered close. For example, 12.1% of respondents in the anchoring group stay within a +/- 1% range of the anchor. Column four gives the p-value of a chi square test checking whether the fraction of respondents within the band is equal across the three groups. The last column gives a rough estimate of the strength of the anchoring effect relative to the status quo bias by calculating: (fraction anchoring - fraction no info) / (fraction status quo - fraction non info).

D.2 Description of Original Survey Instruments

For a short overview of the different surveys, see Table 5.1.

D.2.1 Survey 1

The survey versions for the five groups of Survey 1 are presented one after another: No Info Group, Status Quo Group, Status Quo No Choice Group, Anchoring Group, Captcha Group.

Respondents in the No Info Group received a screen directly asking about their policy bliss points without being given any information on the current status quo policy. See Figure D.1 for a screenshot of the original survey question (in German). The screen said: "The German state has taken several measures to protect the environment. In your opinion, how much money should the German state spend on such measures per capita and year? Please enter your answer into the box provided." On the right-hand side of the box it said "Euros per capita and year".

Figure D.1: Survey 1: No Info Group

Respondents in the Status Quo Group received information about the current status quo policy. See Figure D.2 for screenshots of the original survey questions (in German). The first screen said: "The German state has taken several measures to protect the environment. It spends around 41 Euros per capita and year on such measures. In your opinion, the German state should spend on such measures per capita and year... To answer this question choose one of the three answering options." The answering options were: "more than 41 Euros", "less than 41 Euros", "41 Euros". The order of the different answering options was randomized across respondents. If respondents chose either the "more" or the "less" option, the second screen popped up asking for the exact amount: "In your opinion, how much money should the German state spend on such measures per capita and year? Please enter your answer into the box provided." On the right-hand side of the box it said "Euros per capita and year".

Figure D.2: Survey 1: Status Quo Group

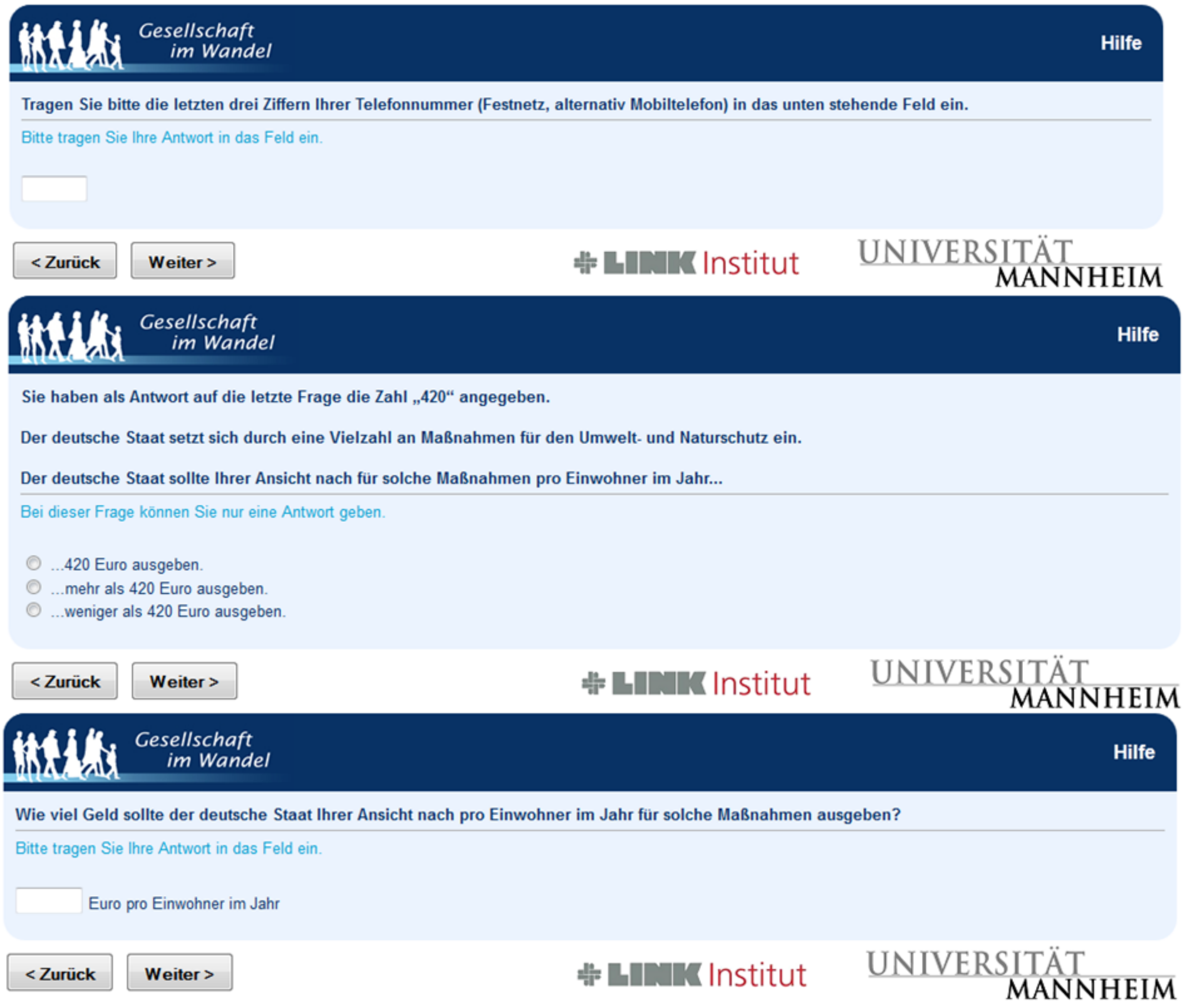
Respondents in the Status Quo No Choice Group received information about the current status quo policy. In contrast to the Status Quo Group, they did not receive a screen with three choice options first, but were directly asked for their exact policy bliss point. See Figure D.3 for a screenshot of the original survey question (in German). The screen said: "The German state has taken several measures to protect the environment. It spends around 41 Euros per capita and year on such measures. In your opinion, how much money should the German state spend on such measures per capita and year? Please enter your answer into the box provided." On the right-hand side of the box it said "Euros per capita and year".

Respondents in the Anchoring Group first were asked to provide some digits of their phone number and then to put their policy bliss point into perspective to it without receiving any information on the current status quo policy. See Figure D.4 for screenshots of the original survey questions (in German). The first screen said: "Please enter the last three digits of your phone number (landline, alternatively mobile) into the box provided below. Please enter your answer into the box provided." On the second screen it said: "You gave number "[the respondent's answer to the previous question was inserted here by the computer]" as response to the previous question. The German state has taken several measures to protect the environment. In your opinion, the German state should spend on such measures per capita and year... To answer this question choose one of the three answering options." The answering options were: "more than "[the respondent's answer to the previous question was inserted here by the computer]"

Figure D.3: Survey 1: Status Quo No Choice Group

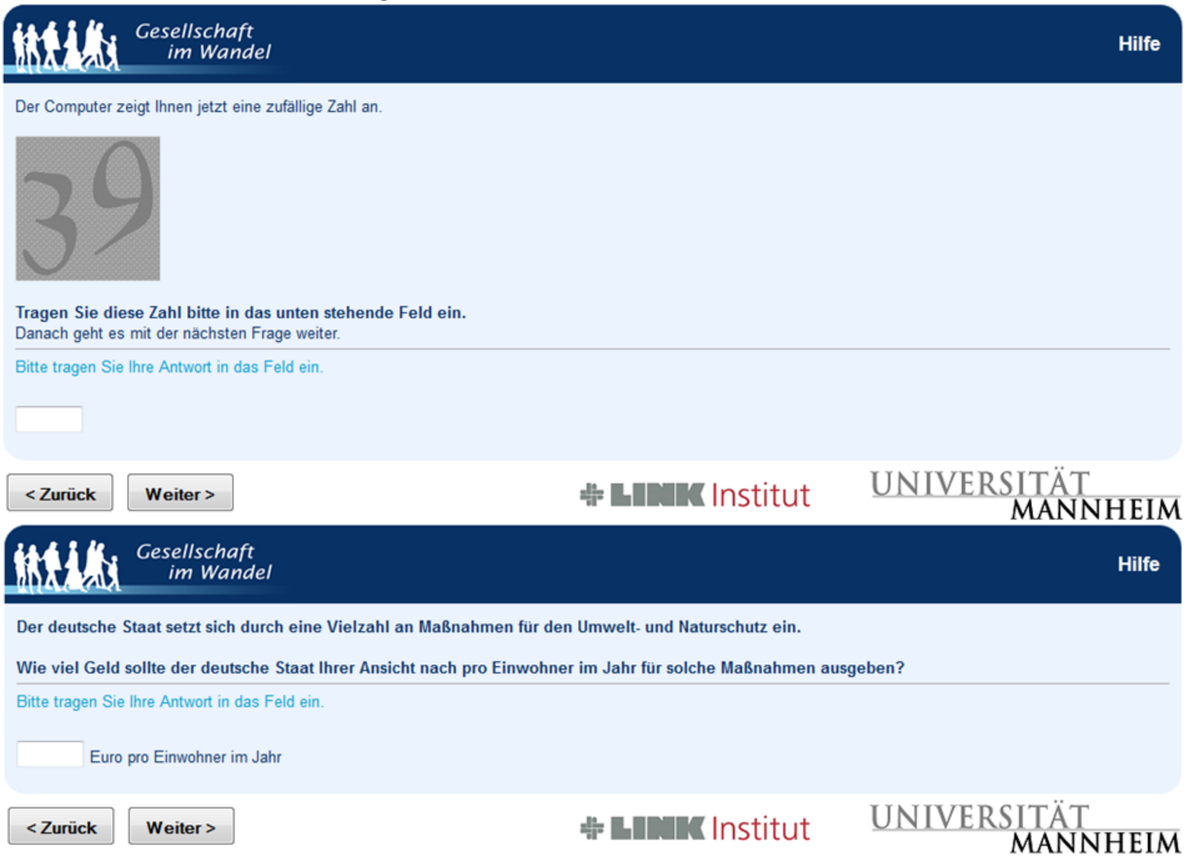
Euros", "less than "[the respondent's answer to the previous question was inserted here by the computer]" Euros", ""[the respondent's answer to the previous question was inserted here by the computer]" Euros". The order of the different answering options was randomized across respondents. If respondents chose either the "more" or the "less" option, a third screen popped up asking for the exact amount: "In your opinion, how much money should the German state spend on such measures per capita and year? Please enter your answer into the box provided." On the right-hand side of the box it said "Euros per capita and year".

Figure D.4: Survey 1: Anchoring Group



Respondents in the Captcha Group were first shown a captcha test and then asked for their policy bliss points. See Figure D.5 for screenshots of the original survey questions (in German). The first screen said: "The computer now shows you a random number." The computer randomly showed one out of three pictures depicting 11, 39, or 390 according to the probability distribution 25%-50%-25%. "Please enter this number into the box provided below. Then proceed to the next question. Please enter your answer into the box provided." The second screen said: "The German state has taken several measures to protect the environment. In your opinion, how much money should the German state spend on such measures per capita and year? Please enter your answer into the box provided." On the right-hand side of the box it said "Euros per capita and year".

Figure D.5: Survey 1: Captcha Group



Some questions unrelated to this project (also by other researchers) chosen by the Board of the GIP followed.

D.2.2 Survey 2

The survey versions for the three groups of Survey 2 are presented one after another: No Info Group, Status Quo Group, Anchoring Group.

Respondents in the No Info Group were first asked about their gender and then about their policy bliss point without receiving any information on the current status quo policy. See Figure D.6 for the original questionnaire (in German). It said:

"Anonymous Questionnaire. Question A: What is your gender? Male or Female? Question B: In Germany, various types of stage performances are publicly subsidized. Among these are performances at playhouses such as plays, operas, children and youth theater, concerts, dances, musicals, operettas, and puppet plays. Besides, stage performances of private theaters, independent orchestras, festivals, and radio orchestras are also publicly subsidized in Germany. In your opinion, how much money should the German state spend on subsidizing such stage performances per capita and year?" Next to the line on which to put the answer it said "Euros per capita and year".

Figure D.6: Survey 2: No Info Group

ANONYMER FRAGEBOGEN

Frage A

Bitte geben Sie Ihr Geschlecht an.

□ männlich □ weiblich

Frage B

In Deutschland werden zahlreiche **Bühnenveranstaltungen** staatlich gefördert. Darunter fallen Veranstaltungen an Schauspielhäusern wie Schauspiel, Oper, Kinder- und Jugendtheater, Konzert, Tanz, Musical, Operette und Figurentheater. Außerdem werden in Deutschland auch Bühnenveranstaltungen von Privattheatern, selbstständigen Kulturorchestern sowie Festspiele und Rundfunkorchester staatlich gefördert.

Wie viel Geld sollte der deutsche Staat Ihrer Ansicht nach pro Einwohner im Jahr für die Förderung solcher Bühnenveranstaltungen ausgeben?

__________ Euro pro Einwohner im Jahr

Respondents in the Status Quo Group were first asked about their gender and then about their policy bliss point while also receiving information on the current status quo policy. See Figure D.7 for the original questionnaire (in German). It said:

"Anonymous Questionnaire. Question A: What is your gender? Male or Female? Question B: In Germany, various types of stage performances are publicly subsidized. Among these are performances at playhouses such as plays, operas, children and youth theater, concerts, dances, musicals, operettas, and puppet plays. Besides, stage performances of private theaters, independent orchestras, festivals, and radio orchestras are also publicly subsidized in Germany. The German state spends about 33 Euros per capita and year on subsidizing such stage performances. In your opinion, the German state should spend per capita and year on subsidizing such stage performances less than 33 Euros, 33 Euros, or more than 33 Euros? Question C: If you answered Question B with "less..." or "more...": In your opinion, how much money should the German state spend on subsidizing such stage performances per capita and year?" Next to the line on which to put the answer it said "Euros per capita and year".

Figure D.7: Survey 2: Status Quo Group

ANONYMER FRAGEBOGEN

Frage A

Bitte geben Sie Ihr Geschlecht an.

□ männlich   □ weiblich

Frage B

In Deutschland werden zahlreiche **Bühnenveranstaltungen** staatlich gefördert. Darunter fallen Veranstaltungen an Schauspielhäusern wie Schauspiel, Oper, Kinder- und Jugendtheater, Konzert, Tanz, Musical, Operette und Figurentheater. Außerdem werden in Deutschland auch Bühnenveranstaltungen von Privattheatern, selbstständigen Kulturorchestern sowie Festspiele und Rundfunkorchester staatlich gefördert.

Der deutsche Staat gibt pro Einwohner im Jahr etwa 33 Euro für die Förderung der eben genannten Arten von Bühnenveranstaltungen aus.

Der deutsche Staat sollte Ihrer Ansicht nach pro Einwohner im Jahr für die Förderung der eben genannten Arten von Bühnenveranstaltungen...

□ weniger als 33 Euro ausgeben.

□ genau 33 Euro ausgeben.

□ mehr als 33 Euro ausgeben.

Frage C

Falls Sie Frage B mit „weniger...“ oder „mehr...“ beantwortet haben:

Wie viel Geld sollte der deutsche Staat Ihrer Ansicht nach pro Einwohner im Jahr für die Förderung solcher Bühnenveranstaltungen ausgeben?

__________ Euro pro Einwohner im Jahr

Respondents in the Anchoring Group were first asked about their gender, then to give digits of a phone number, and then about their policy bliss point without receiving any information on the current status quo policy. See Figure D.8 for the original questionnaire (in German). It said:

"Anonymous Questionnaire. Question A: What is your gender? Male or Female? Question B: Please enter the last two digits of your phone number (landline, alternatively mobile) into the box provided below. Question C: In Germany, various types of stage performances are publicly subsidized. Among these are performances at playhouses such as plays, operas, children and youth theater, concerts, dances, musicals, operettas, and puppet plays. Besides, stage performances of private theaters, independent orchestras, festivals, and radio orchestras are also publicly subsidized in Germany. Now consider the two-digit number you put into the box in Question B. In your opinion, the German state should spend per capita and year on subsidizing such stage performances less than that number in Euros, exactly that number in Euros, or more than that number in Euros? Question C: If you answered Question C with "less..." or "more...": In your opinion, how much money should the German state spend on subsidizing such stage performances per capita and year?" Next to the line on which to put the answer it said "Euros per capita and year".

Figure D.8: Survey 2: Anchoring Group

ANONYMER FRAGEBOGEN

Frage A

Bitte geben Sie Ihr Geschlecht an.

□ männlich □ weiblich

Frage B

Tragen Sie bitte die letzten beiden Ziffern Ihrer Telefonnummer (Festnetz, alternativ Mobiltelefon) in das unten stehende umrahmte Feld ein:

Frage C

In Deutschland werden zahlreiche **Bühnenveranstaltungen** staatlich gefördert. Darunter fallen Veranstaltungen an Schauspielhäusern wie Schauspiel, Oper, Kinder- und Jugendtheater, Konzert, Tanz, Musical, Operette und Figurentheater. Außerdem werden in Deutschland auch Bühnenveranstaltungen von Privattheatern, selbstständigen Kulturorchestern sowie Festspiele und Rundfunkorchester staatlich gefördert.

Betrachten Sie nun die zweistellige Zahl, die Sie in Frage B in das umrahmte Feld eingetragen haben. Der deutsche Staat sollte Ihrer Ansicht nach pro Einwohner im Jahr für die Förderung der eben genannten Arten von Bühnenveranstaltungen…

- □ weniger als diese Zahl in Euro ausgeben.
- □ genau diese Zahl in Euro ausgeben.
- □ mehr als diese Zahl in Euro ausgeben.

Frage D

Falls Sie Frage C mit „weniger…" oder „mehr…" beantwortet haben:

Wie viel Geld sollte der deutsche Staat Ihrer Ansicht nach pro Einwohner im Jahr für die Förderung solcher Bühnenveranstaltungen ausgeben?

__________ Euro pro Einwohner im Jahr

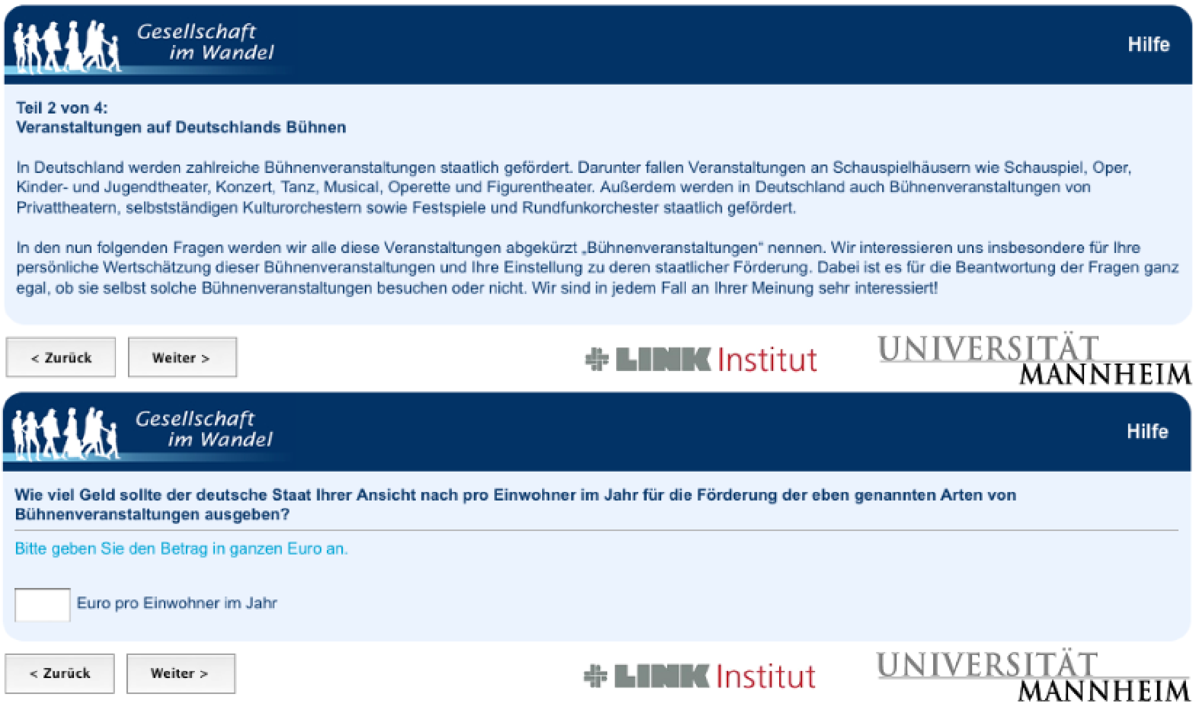
In all three versions some questions unrelated to this project followed.

D.2.3 Survey 3

The survey versions for the four groups of Survey 3 are presented one after another: No Info Group, Status Quo Group, Anchoring Group (i.e. phone anchor), Basic Anchoring Group (i.e. captcha anchor). The introductory screen was the same across all four versions (first screen in each of Figures D.9, D.10, D.11, and D.12). It said: "Part 2 of 4 [the other survey parts belonged to other research projects by other researchers]: Performances on German Stages. In Germany, various types of stage performances are publicly subsidized. Among these are performances at playhouses such as plays, operas, children and youth theater, concerts, dances, musicals, operettas, and puppet plays. Besides, stage performances of private theaters, independent orchestras, festivals, and radio orchestras are also publicly subsidized in Germany. In subsequent questions all these performances will shortly be called “stage performance”. We are particularly interested in your personal valuation of these stage performances and your attitude towards them being publicly subsidized. It is totally irrelevant for answering the questions if you yourself go to such stage performances are not. We are interested in your opinion in any case!”

Respondents in the No Info Group were directly asked about their policy bliss point without being given any information on the current status quo policy. See Figure D.9 for screenshots of the original survey questions (in German). The second screen said: “In your opinion, how much money should the German state spend on subsidizing such stage performances per capita and year? Please give the amount in whole Euros.” On the right-hand side of the box it said “Euros per capita and year”.

Figure D.9: Survey 3: No Info Group


Gesellschaft im Wandel

Hilfe

Teil 2 von 4:
Veranstaltungen auf Deutschlands Bühnen

In Deutschland werden zahlreiche Bühnenveranstaltungen staatlich gefördert. Darunter fallen Veranstaltungen an Schauspielhäusern wie Schauspiel, Oper, Kinder- und Jugendtheater, Konzert, Tanz, Musical, Operette und Figurentheater. Außerdem werden in Deutschland auch Bühnenveranstaltungen von Privattheatern, selbstständigen Kulturorchestern sowie Festspiele und Rundfunkorchester staatlich gefördert.

In den nun folgenden Fragen werden wir alle diese Veranstaltungen abgekürzt „Bühnenveranstaltungen“ nennen. Wir interessieren uns insbesondere für Ihre persönliche Wertschätzung dieser Bühnenveranstaltungen und Ihre Einstellung zu deren staatlicher Förderung. Dabei ist es für die Beantwortung der Fragen ganz egal, ob sie selbst solche Bühnenveranstaltungen besuchen oder nicht. Wir sind in jedem Fall an Ihrer Meinung sehr interessiert!

< Zurück | Weiter >

LINK Institut

UNIVERSITÄT MANNHEIM

Gesellschaft im Wandel

Hilfe

Wie viel Geld sollte der deutsche Staat Ihrer Ansicht nach pro Einwohner im Jahr für die Förderung der eben genannten Arten von Bühnenveranstaltungen ausgeben?

Bitte geben Sie den Betrag in ganzen Euro an.

Euro pro Einwohner im Jahr

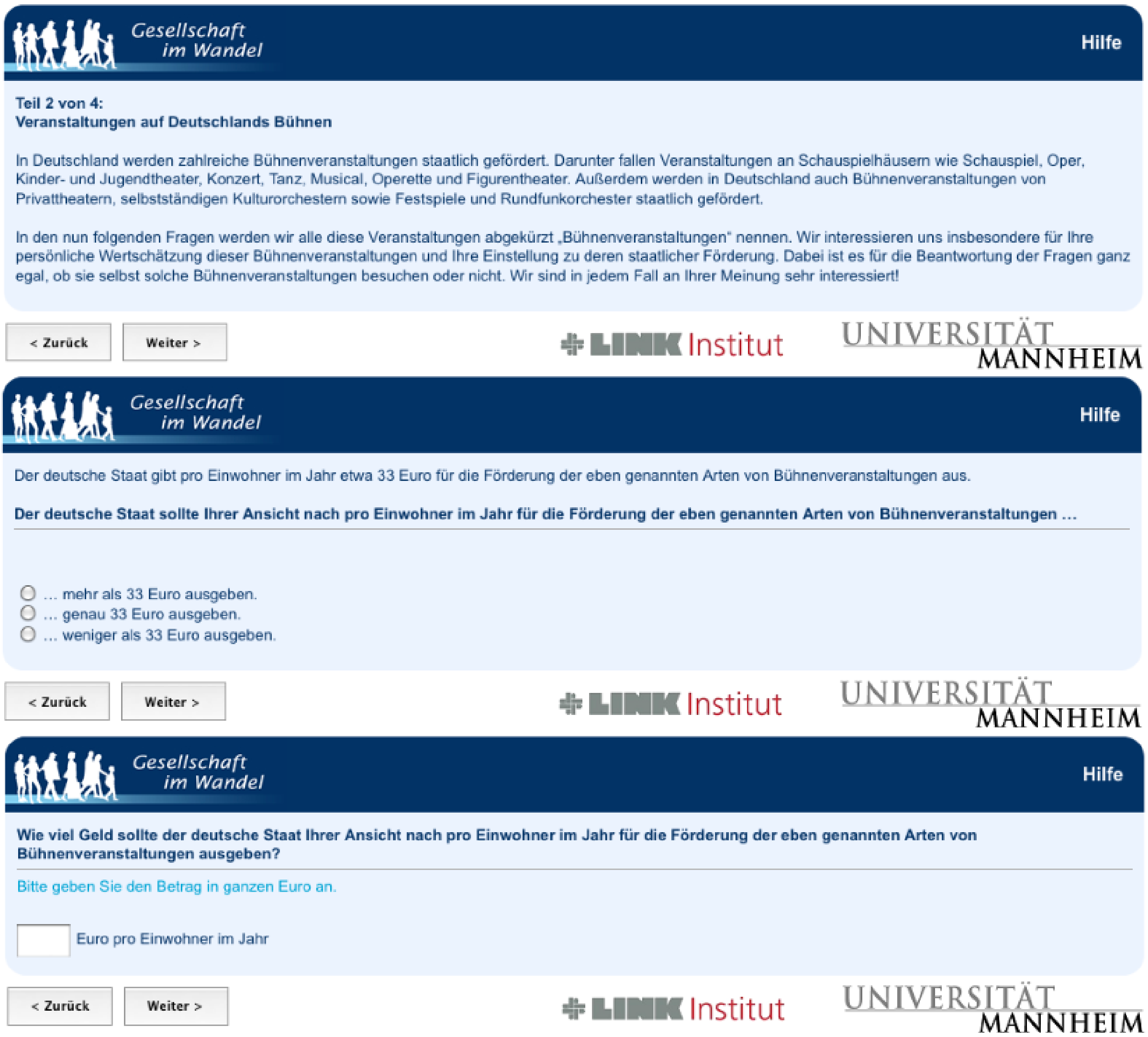
< Zurück | Weiter >

LINK Institut

UNIVERSITÄT MANNHEIM

Respondents in the Status Quo Group received information on the current status quo policy and were then asked about their policy bliss point. See Figure D.10 for screenshots of the original survey questions (in German). The second screen said: “The German state spends about 33 Euros per capita and year on subsidizing such stage performances. In your opinion, the German state should spend per capita and year on such stage performances more than 33 Euros, exactly 33 Euros, or less than 33 Euros?” The order of the different answering options was randomized across respondents. If the respondent answered “more” or “less”, a third screen asking for the exact amount popped up: “In your opinion, how much money should the German state spend on subsidizing such stage performances per capita and year? Please give the amount in whole Euros.” On the right-hand side of the box it said “Euros per capita and year”.

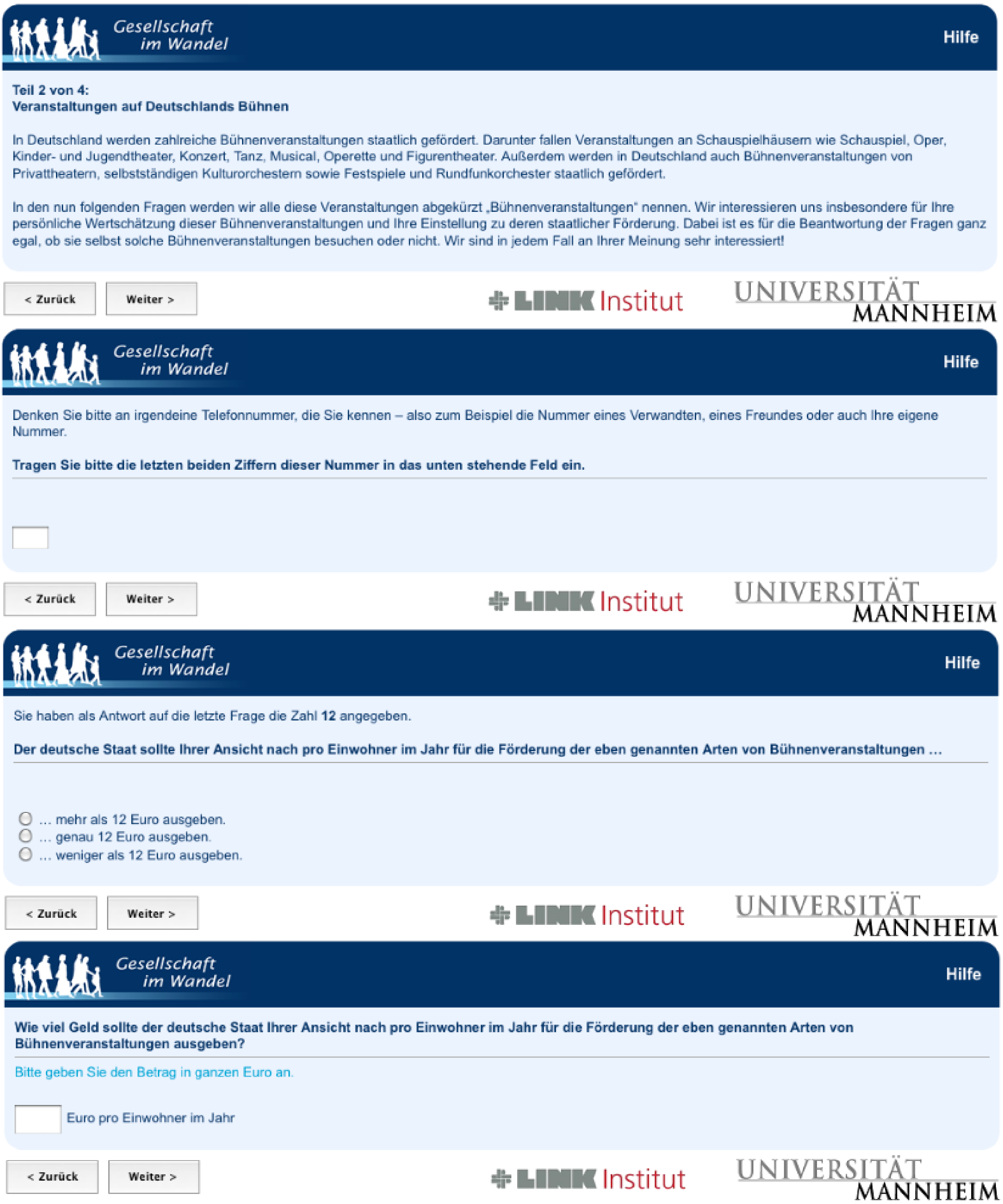
Figure D.10: Survey 3: Status Quo Group



Respondents in the Anchoring Group were first asked about their phone number and then to relate it to their policy bliss point without being given any information on the current status quo policy. See Figure D.11 for screenshots of the original survey questions (in German). The second screen said: "Consider any phone number you know - for example, the number of a relative, a friend, or also your own number. Please enter the last two digits of that number into the box provided below." The third screen continued: "You gave the number [the respondent's answer to the previous question was inserted in bold here by the computer] as response to the previous question. In your opinion, the German state should spend per capita and year on subsidizing such stage performances "more than [the respondent's answer to the previous question was inserted here by the computer] Euros, exactly [the respondent's answer to the previous question was inserted here by the computer] Euros", or less than [the respondent's answer to the previous question was inserted here by the computer] Euros." The order of the different answering options was randomized across respondents. If respondents chose either the "more" or the "less" option, a fourth screen popped up asking for the exact amount: "In your opinion, how much money should the German state spend per capita and year on subsidizing such stage performances?

Please give the amount in whole Euros.” On the right-hand side of the box it said “Euros per capita and year”.

Figure D.11: Survey 3: Anchoring Group



Gesellschaft im Wandel

Hilfe

Teil 2 von 4:
Veranstaltungen auf Deutschlands Bühnen

In Deutschland werden zahlreiche Bühnenveranstaltungen staatlich gefördert. Darunter fallen Veranstaltungen an Schauspielhäusern wie Schauspiel, Oper, Kinder- und Jugendtheater, Konzert, Tanz, Musical, Operette und Figurentheater. Außerdem werden in Deutschland auch Bühnenveranstaltungen von Privattheatern, selbstständigen Kulturorchestern sowie Festspiele und Rundfunkorchester staatlich gefördert.

In den nun folgenden Fragen werden wir alle diese Veranstaltungen abgekürzt „Bühnenveranstaltungen" nennen. Wir interessieren uns insbesondere für Ihre persönliche Wertschätzung dieser Bühnenveranstaltungen und Ihre Einstellung zu deren staatlicher Förderung. Dabei ist es für die Beantwortung der Fragen ganz egal, ob sie selbst solche Bühnenveranstaltungen besuchen oder nicht. Wir sind in jedem Fall an Ihrer Meinung sehr interessiert!

< Zurück | Weiter >

LINK Institut — UNIVERSITÄT MANNHEIM

Gesellschaft im Wandel

Hilfe

Denken Sie bitte an irgendeine Telefonnummer, die Sie kennen – also zum Beispiel die Nummer eines Verwandten, eines Freundes oder auch Ihre eigene Nummer.

Tragen Sie bitte die letzten beiden Ziffern dieser Nummer in das unten stehende Feld ein.

< Zurück | Weiter >

LINK Institut — UNIVERSITÄT MANNHEIM

Gesellschaft im Wandel

Hilfe

Sie haben als Antwort auf die letzte Frage die Zahl **12** angegeben.

Der deutsche Staat sollte Ihrer Ansicht nach pro Einwohner im Jahr für die Förderung der eben genannten Arten von Bühnenveranstaltungen …

- … mehr als 12 Euro ausgeben.
- … genau 12 Euro ausgeben.
- … weniger als 12 Euro ausgeben.

< Zurück | Weiter >

LINK Institut — UNIVERSITÄT MANNHEIM

Gesellschaft im Wandel

Hilfe

Wie viel Geld sollte der deutsche Staat Ihrer Ansicht nach pro Einwohner im Jahr für die Förderung der eben genannten Arten von Bühnenveranstaltungen ausgeben?

Bitte geben Sie den Betrag in ganzen Euro an.

Euro pro Einwohner im Jahr

< Zurück | Weiter >

LINK Institut — UNIVERSITÄT MANNHEIM

Respondents in the Captcha Group were first asked to complete a captcha question and then about their policy bliss point without being given any information on the current status quo policy. See Figure D.12 for screenshots of the original survey questions (in German). The second screen said: “The computer now shows you a random number.” The computer randomly showed one out of three pictures depicting 33, 66, or 99 according to the probability distribution 66.66%-16.66%-16.66%. “Please enter this number into the box provided below. Then continue to the next question.” The third screen said: “In your opinion, how much money should the German state spend per capita and year on subsidizing such stage performances?” On the right-hand side of the box it said “Euros per capita and year”.

Figure D.12: Survey 3: Captcha Group


Gesellschaft im Wandel

Hilfe

Teil 2 von 4:
Veranstaltungen auf Deutschlands Bühnen

In Deutschland werden zahlreiche Bühnenveranstaltungen staatlich gefördert. Darunter fallen Veranstaltungen an Schauspielhäusern wie Schauspiel, Oper, Kinder- und Jugendtheater, Konzert, Tanz, Musical, Operette und Figurentheater. Außerdem werden in Deutschland auch Bühnenveranstaltungen von Privattheatern, selbstständigen Kulturorchestern sowie Festspiele und Rundfunkorchester staatlich gefördert.

In den nun folgenden Fragen werden wir alle diese Veranstaltungen abgekürzt „Bühnenveranstaltungen“ nennen. Wir interessieren uns insbesondere für Ihre persönliche Wertschätzung dieser Bühnenveranstaltungen und Ihre Einstellung zu deren staatlicher Förderung. Dabei ist es für die Beantwortung der Fragen ganz egal, ob sie selbst solche Bühnenveranstaltungen besuchen oder nicht. Wir sind in jedem Fall an Ihrer Meinung sehr interessiert!

< Zurück | Weiter >

LINK Institut UNIVERSITÄT MANNHEIM

Gesellschaft im Wandel

Hilfe

Der Computer zeigt Ihnen jetzt eine zufällige Zahl an.

33

Tragen Sie bitte diese Zahl in das unten stehende Feld ein.
Danach geht es mit der nächsten Frage weiter.

< Zurück | Weiter >

LINK Institut UNIVERSITÄT MANNHEIM

Gesellschaft im Wandel

Hilfe

Wie viel Geld sollte der deutsche Staat Ihrer Ansicht nach pro Einwohner im Jahr für die Förderung der eben genannten Arten von Bühnenveranstaltungen ausgeben?

Euro pro Einwohner im Jahr

< Zurück | Weiter >

LINK Institut UNIVERSITÄT MANNHEIM

Some questions unrelated to this project (also by other researchers) chosen by the Board of the GIP followed.

Bibliography

Abgeordnetenwatch.de (2011). www.abgeordnetenwatch.de. (last access: April 7th, 2014).

Acemoglu, D. (2005). Constitutions, Politics, and Economics: A Review Essay on Persson and Tabellini's the Economic Effects of Constitutions. *Journal of Economic Literature 43*(4), 1025–1048.

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

Alesina, A., P. Giuliano, and N. Nunn (2014). On the Origins of Gender Roles: Women and the Plough. *Quarterly Journal of Economics*, forthcoming.

Alesina, A. and H. Rosenthal (1995). *Partisan Politics, Divided Government, and the Economy.* Cambridge University Press, New York.

Alesina, A. and H. Rosenthal (1996). A Theory of Divided Government. *Econometrica 64*(6), 1311–1341.

Alesina, A. and H. Rosenthal (2000). Polarized platforms and moderate policies with checks and balances. *Journal of Public Economics 75*(1), 1–20.

Allen, R. and J. Clark (1981). State Policy Adoption and Innovation: Lobbying and Education. *State and Local Government Review 13*(1), 18–25.

Alpizar, F., F. Carlsson, and O. Johansson-Stenman (2008). Anonymity, reciprocity, and conformity: Evidence from voluntary contributions to a national park in Costa Rica. *Journal of Public Economics 92*(5–6), 1047–1060.

Alt, J. E. and R. C. Lowry (1994). Divided Government, Fiscal Institutions, and Budget Deficits: Evidence from the States. *American Political Science Review 88*(4), 811–828.

Andersen, A. L., D. D. Lassen, and L. H. W. Nielsen (2010). Fiscal Governance and Electoral Accountability: Evidence from Late Budgets. *Paper presented at the 2010 Meeting of the American Political Science Association*.

Andersen, A. L., D. D. Lassen, and L. H. W. Nielsen (2012). Late Budgets. *American Economic Journal: Economic Policy 4*(4), 1–40.

Anderson, C. J. (2003). The psychology of doing nothing: Forms of decision avoidance result from reason and emotion. *Psychological Bulletin Vol 129(1)*, 139–166.

Ariely, D., G. Loewenstein, and D. Prelec (2003). "Coherent Arbitrariness": Stable Demand Curves without Stable Preferences. *Quarterly Journal of Economics 118*(1), 73–105.

Arnold, F. (2013). German MPs' Outside Jobs and Their Repercussions on Parliamentary Effort. *DIW Discussion Paper 1340*.

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.

Bateman, I., A. Munro, B. Rhodes, C. Starmer, and R. Sugden (1997). A Test of the Theory of Reference-Dependent Preferences. *Quarterly Journal of Economics 112*(2), 479–505.

Becker, J., A. Peichl, and J. Rincke (2009). Politicians' outside earnings and electoral competition. *Public Choice 140*(3-4), 379–394.

Beggs, A. and K. Graddy (2009). Anchoring Effects: Evidence from Art Auctions. *American Economic Review 99*(3), 1027–1039.

Bergman, O., T. Ellingsen, M. Johannesson, and C. Svensson (2010). Anchoring and cognitive ability. *Economics Letters 107*(1), 66–68.

Berliner Zeitung (2012, October 9th). http://www.bz-berlin.de/aktuell/deutschland/steinbrueck-schwaenzte-bundestag-fuer-reden-article1558723.html.

Bernecker, A. (2011). Electoral Competition and Shirking in Politics: Micro-Evidence from Germany. *University of Mannheim Unpublished Master Thesis.*

Bernecker, A. (2013). Do Politicians Shirk when Reelection Is Certain? Evidence from the German Parliament. *University of Mannheim Economics Department Working Paper ECON 13-09.*

Bernecker, A. (2014). Divided We Reform? Evidence from US Welfare Policies. *CESifo Working Paper No. 4564*.

Bernecker, A. and C. Gathmann (2013). Trial and Error? Policy Experimentation during the US Welfare Reform. *University of Mannheim Economics Department Working Paper ECON 13-05*.

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.

Bertrand, M., E. Duflo, and S. Mullainathan (2004). How Much Should We Trust Differences-in-Differences Estimates? *Quarterly Journal of Economics 119*(1), 249–275.

Besley, T. (2006). *Principled agents? : The political economy of good government*. Oxford; New York: 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.

Besley, T., O. Folke, T. Persson, and J. Rickne (2013). Gender Quotas and the Crisis of the Mediocre Man: Theory and Evidence from Sweden. *Working Paper*.

Besley, T. and V. Larcinese (2011). Working or shirking? Expenses and attendance in the UK Parliament. *Public Choice 146*(3), 291–317.

Besley, T., T. Persson, and D. M. Sturm (2010). Political Competition, Policy and Growth: Theory and Evidence from the US. *Review of Economic Studies 77*(4), 1329–1352.

Besley, T. and I. Preston (2007). Electoral Bias and Policy Choice: Theory and Evidence. *Quarterly Journal of Economics 122*(4), 1473–1510.

Binder, S. A. (1999). The Dynamics of Legislative Gridlock, 1947-96. *American Political Science Review 93*(3), 519–533.

Binder, S. A. (2003). *Stalemate: Causes and Consequences of Legislative Gridlock*. Washington, DC: Brookings Institution Press.

Binder, S. A. (2011). Legislative Productivity and Gridlock. In E. Schickler (Ed.), *The Oxford Handbook of the American Congress*, Chaper 28. Oxford University Press.

Bitler, M. and H. Hoynes (2010). The State of the Safety Net in the Post-Welfare Reform Era. *Brookings Papers on Economic Activity Fall*, 71–127.

Bjørnskov, C. and N. Potrafke (2013). The size and scope of government in the us states: Does party ideology matter? *International Tax and Public Finance 20 (4)*, 687–714.

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

Boehmke, F. J. and P. Skinner (2012). State Policy Innovativeness Revisited. *State Politics & Policy Quarterly 12*(3), 303–329.

Bowling, C. J. and M. R. Ferguson (2001). Divided Government, Interest Representation, and Policy Differences: Competing Explanations of Gridlock in the Fifty States. *Journal of Politics 63*, 182–206.

Brewer, N. T. and G. B. Chapman (2002). The fragile basic anchoring effect. *Journal of Behavioral Decision Making 15*(1), 65–77.

Brueckner, J. K. (2000). Welfare Reform and the Race to the Bottom: Theory and Evidence. *Southern Economic Journal 66*(3), 505–525.

Bundestag (2012). www.bundestag.de. (last access: April 7th, 2014).

Bundeswahlleiter (2012). www.bundeswahlleiter.de. (last access: April 7th, 2014).

Carson, R. and T. Groves (2007). Incentive and informational properties of preference questions. *Environmental and Resource Economics 37*, 181–210.

Castanheira, M., G. Nicodeme, and P. Profeta (2012). On the political economics of tax reforms: survey and empirical assessment. *International Tax and Public Finance 19*(4), 598–624.

Caughey, D. and J. S. Sekhon (2011). Elections and the Regression Discontinuity Design: Lessons from Close U.S. House Races, 1942–2008. *Political Analysis 19*(4), 385–408.

Center for American Women and Politics (2012). Facts on Women Candidates. http://www.cawp.rutgers.edu/fast_facts/. (last access: February 7th, 2014).

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

Chapman, G. B. and E. J. Johnson (1999). Anchoring, Activation, and the Construction of Values. *Organizational Behavior and Human Decision Processes 79*(2), 115–153.

Chari, V. V., L. E. Jones, and R. Marimon (1997). The Economics of Split-Ticket Voting in Representative Democracies. *American Economic Review 87*(5), 957–976.

Chattopadhyay, R. and E. Duflo (2004). Women as Policy Makers: Evidence from a Randomized Policy Experiment in India. *Econometrica 72*(5), 1409–1443.

Cnudde, C. F. and D. J. McCrone (1969). Party Competition and Welfare Policies in the American States. *American Political Science Review 63*, 858–66.

Coate, S. and B. Knight (2007). Socially Optimal Districting: A Theoretical and Empirical Exploration. *Quarterly Journal of Economics 122*(4), 1409–1471.

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

Council of State Governments (2012). *Book of the States.* Washington D.C.: Council of State Governments. Also earlier years.

Critcher, C. R. and T. Gilovich (2008). Incidental environmental anchors. *Journal of Behavioral Decision Making 21*(3), 241–251.

Crouse, G. (1999). Implementation of Major Changes to Welfare Policies, 1992-1998. http://www.aspe.hhs.gov/hsp/Waiver-Policies99/policy_CEA.htm (last access: February 7th, 2014).

Cruces, G., R. Perez-Truglia, and M. Tetaz (2013). Biased perceptions of income distribution and preferences for redistribution: Evidence from a survey experiment. *Journal of Public Economics 98*(0), 100–112.

Cukierman, A. and M. Tommasi (1998). When Does It Take a Nixon to Go to China? *American Economic Review 88*(1), 180–197.

Cummings, R. G., S. Elliott, G. W. Harrison, and J. Murphy (1997). Are Hypothetical Referenda Incentive Compatible? *Journal of Political Economy 105*(3), 609–621.

Cummings, R. G., G. W. Harrison, and E. E. Rutström (1995). Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive-Compatible? *American Economic Review 85*(1), 260–266.

Cummings, R. G. and L. O. Taylor (1999). Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method. *American Economic Review 89*(3), 649–665.

Cutler, L. N. (1988). Some Reflections About Divided Government. *Presidential Studies Quarterly 18*(3), 485–492.

Dal Bó, E., P. Dal Bó, and J. Snyder (2009). Political Dynasties. *Review of Economic Studies 76*(1), 115–142.

Dell, M. (2010). The Persistent Effects of Peru's Mining Mita. *Econometrica 78*(6), 1863–1903.

Dewatripont, M. and G. Roland (1995). The Design of Reform Packages under Uncertainty. *American Economic Review 85*(5), 1207–1223.

Economist, T. (2011a). Austerity in Southern Europe: Spain's cry of pain. *May 28th*, p. 13.

Economist, T. (2011b). Constitutional reform in Britain: Yes or No? *April 30th*, p. 13.

Economist, T. (2011c). Election boundaries: No more packing or cracking. *June 18th*, p. 49.

Eggers, A. C., O. Folke, A. Fowler, J. Hainmueller, A. B. Hall, and J. M. Snyder (May 2013). On The Validity Of The Regression Discontinuity Design For Estimating Electoral Effects: New Evidence From Over 40,000 Close Races. *Working Paper*.

Eidelman, S. and C. Crandall (2009). A Psychological Advantage for the Status Quo. In J. T. Jost, A. C. Kay, and H. Thorisdottir (Eds.), *Social and Psychological Bases of Ideology and System Justification*, pp. 85–106. New York, NY: Oxford University Press.

Engelmann, D. and G. Hollard (2010). Reconsidering the Effect of Market Experience on the "Endowment Effect". *Econometrica 78*(6), 2005–2019.

Epley, N. (2004). A Tale of Tuned Decks? Anchoring as Adjustment and Anchoring as Activation. In D. J. Koehler and N. Harvey (Eds.), *Blackwell Handbook of Judgement and Decision Making*, pp. 240–56. Malden, MA: Blackwell Pub.

Epley, N. and T. Gilovich (2001). Putting Adjustment Back in the Anchoring and Adjustment Heuristic: Differential Processing of Self-Generated and Experimenter-Provided Anchors. *Psychological Science 12*(5), 391–396.

Epley, N. and T. Gilovich (2006). The Anchoring-and-Adjustment Heuristic. *Psychological Science 17*(4), 311–318.

Epley, N. and T. Gilovich (2010). Anchoring unbound. *Journal of Consumer Psychology 20*(1), 20–24.

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

Fehr, E. and J.-R. Tyran (2008). Limited Rationality and Strategic Interaction: The Impact of the Strategic Environment on Nominal Inertia. *Econometrica 76*(2), 353–394.

Fellowes, M. C. and G. Rowe (2004). Politics and the New American Welfare States. *American Journal of Political Science 48*(2), 362–373.

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: Evidence from the Audits of Local Governments. *American Economic Review 101*(4), 1274–1311.

Ferreira, F. and J. Gyourko (2009). Do Political Parties Matter? Evidence from U.S. Cities. *Quarterly Journal of Economics 124*(1), 399–422.

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

Fiorina, M. (1996). *Divided Government* (second ed.). Allyn & Bacon.

Fisman, R., N. A. Harmon, E. Kamenica, and I. Munk (2012). Labor Supply of Politicians. Working Paper 17726, National Bureau of Economic Research.

Folke, O. and J. M. Snyder (2012). Gubernatorial Midterm Slumps. *American Journal of Political Science 56*(4), 931–948.

Fox, J. and R. V. Weelden (2010). Partisanship and the effectiveness of oversight. *Journal of Public Economics 94*(9–10), 674–687.

Francis, R. (1998). Predictions, Patterns, and Policymaking: A Regional Study of Devolution. *Publius: The Journal of Federalism 28*, 143–160.

Fu, Q. and M. Li (2014). Reputation-concerned policy makers and institutional status quo bias. *Journal of Public Economics 110*(0), 15–25.

Fudenberg, D., D. K. Levine, and Z. Maniadis (2012). On the Robustness of Anchoring Effects in WTP and WTA Experiments. *American Economic Journal: Microeconomics 4*(2), 131–45.

Fujiwara, I., H. Ichiue, Y. Nakazono, and Y. Shigemi (2013). Financial markets forecasts revisited: Are they rational, stubborn or jumpy? *Economics Letters 118*(3), 526–530.

Funk, P. and C. Gathmann (2014). Gender Gaps in Policy Preferences: Evidence from Direct Democracy in Switzerland. *Economic Policy*, forthcoming.

Furnham, A. and H. C. Boo (2011). A literature review of the anchoring effect. *Journal of Socio-Economics 40*(1), 35–42.

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.

Gagliarducci, S., T. Nannicini, and P. Naticchioni (2010). Moonlighting politicians. *Journal of Public Economics 94*(9-10), 688–699.

Gagliarducci, S., T. Nannicini, and P. Naticchioni (2011). Electoral Rules and Politicians' Behavior: A Micro Test. *American Economic Journal: Economic Policy 3*(3), 144–74.

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

Geys, B. and K. Mause (2012). Delegation, Accountability and Legislator Moonlighting: Agency Problems in Germany. *German Politics 21:3*, 255–273.

Geys, B. and K. Mause (2013). Moonlighting Politicians: A Survey and Research Agenda. *Journal of Legislative Studies 19*(1), 76–97.

Gray, V. and R. L. Hanson (Eds.) (2008). *Politics in the American States: A Comparative Analysis* (ninth ed.). CQ Press, Washington D.C.

Green, D., K. E. Jacowitz, D. Kahneman, and D. McFadden (1998). Referendum contingent valuation, anchoring, and willingness to pay for public goods. *Resource and Energy Economics 20*(2), 85–116.

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

Hamburger Abendblatt (2011, September 29th). http://www.abendblatt.de/politik/article2043500/Panorama-Koch-Mehrin-schwaenzt-alle-Sitzungen.html.

Han, W.-J., C. Ruhm, and J. Waldfogel (2009). Parental leave policies and parents' employment and leave-taking. *Journal of Policy Analysis and Management 28*(1), 29–54.

Handelsblatt (2011). Komplexe rechnerei. *June 15th*, p. 14.

Harrington, Joseph E., J. (1993). Economic Policy, Economic Performance, and Elections. *American Economic Review 83*(1), 27–42.

Hartman, R. S., M. J. Doane, and C.-K. Woo (1991). Consumer Rationality and the Status Quo. *Quarterly Journal of Economics 106*(1), 141–162.

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*. Washington DC: The Brookings Institution.

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.

Howitt, P. and R. Wintrobe (1995). The political economy of inaction. *Journal of Public Economics 56*(3), 329–353.

Hoxby, C. M. (2000). The Effects of Class Size on Student Achievement: New Evidence from Population Variation. *Quarterly Journal of Economics 115*(4), 1239–1285.

Imbens, G. and T. Zajonc (2011). Regression Discontinuity Design with Multiple Forcing Variables. *Working Paper*.

Imbens, G. W. and T. Lemieux (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics 142*(2), 615–635.

Jacobson, G. C. (1990). *The electoral origins of divided government: Competition in US House elections, 1946-1988*. Boulder: Westview Press.

Jacowitz, K. E. and D. Kahneman (1995). Measures of Anchoring in Estimation Tasks. *Personality and Social Psychology Bulletin 21*(11), 1161–1166.

Jain, S. and S. W. Mukand (2003). Redistributive Promises and the Adoption of Economic Reform. *American Economic Review 93*(1), 256–264.

Johnson, E. J. and D. Goldstein (2003). Do Defaults Save Lives? *Science 302*(5649), 1338–1339.

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.

Kauder, B. and N. Potrafke (2014). Outside earnings, attendance, and activity: Do German MPs meet their obligations? *mimeo*.

Klarner, C. (2003). The Measurement of the Partisan Balance of State Government. *State Politics & Policy Quarterly 3*(3), 309–319.

Klarner, C. E., J. H. Phillips, and M. Muckler (2012). Overcoming Fiscal Gridlock: Institutions and Budget Bargaining. *Journal of Politics 74*(4), 992–1009.

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. (2008). Randomized experiments from non-random selection in U.S. house elections. *Journal of Econometrics 142*(2), 675–697.

Lee, D. S. and T. Lemieux (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature 48*(2), 281–355.

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, February). Dave Leip's Atlas of U.S. Presidential Elections. http://www.uselectionatlas.org. (last access: February 7th, 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.

Liebschutz, S. F. (Ed.) (2000). *Managing Welfare Reform in Five States.* Rockefeller Institute Press.

List, J. A. (2001). Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auctions for Sportscards. *American Economic Review 91*(5), 1498–1507.

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.

Lott, John R, J. (1987). Political cheating. *Public Choice 52*(2), 169–186.

Lott, J. R. and L. W. Kenny (1999). Did Women's Suffrage Change the Size and Scope of Government? *Journal of Political Economy 107*(6), 1163–1198.

Madrian, B. C. and D. F. Shea (2001). The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior. *Quarterly Journal of Economics 116*(4), 1149–1187.

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

Maniadis, Z., F. Tufano, and J. A. List (2014). One Swallow Doesn't Make a Summer: New Evidence on Anchoring Effects. *American Economic Review 104*(1), 277–90.

Mayhew, D. R. (2005). *Divided We Govern. Party Control, Lawmaking, and Institutions, 1946-2002* (second ed.). Yale University Press, New Haven and London.

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

Messer, K. D., H. Zarghamee, H. M. Kaiser, and W. D. Schulze (2007). New hope for the voluntary contributions mechanism: The effects of context. *Journal of Public Economics 91*(9), 1783–1799.

Mocan, N. and D. T. Altindag (2013). Salaries and Work Effort: An Analysis of the European Union Parliamentarians. *Economic Journal 123*(573), 1130–1167.

Moen, E. R. and C. Riis (2010). Policy Reversal. *American Economic Review 100*(3), 1261–68.

Moffitt, R. A. (2001). The effect of pre-PRWORA waivers on AFDC caseloads and female earnings, income, and labor force behavior. In S. Danziger (Ed.), *Economic Conditions and Welfare Reform.* W.E. Upjohn Institute.

Moffitt, R. A. (2002a). Welfare Benefits File. http://econ.jhu.edu/directory/robert-a-moffitt/. (last access: February 7th, 2014).

Moffitt, R. A. (2002b). 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.

Mussweiler, T. and B. Englich (2005). Subliminal anchoring: Judgmental consequences and underlying mechanisms. *Organizational Behavior and Human Decision Processes 98*(2), 133–143.

Mussweiler, T. and F. Strack (2001). The Semantics of Anchoring. *Organizational Behavior and Human Decision Processes 86*(2), 234–255.

Nannicini, T., A. Stella, G. Tabellini, and U. Troiano (2013). Social Capital and Political Accountability. *American Economic Journal: Economic Policy 5*(2), 222–50.

Oates (1999). An Essay on Fiscal Federalism. *Journal of Economic Literature 37*(3), 1120–1149.

Papay, J. P., J. B. Willett, and R. J. Murnane (2011). Extending the regression-discontinuity approach to multiple assignment variables. *Journal of Econometrics 161*(2), 203–207.

Peichl, A., N. Pestel, and S. Siegloch (2013). The politicians' wage gap: insights from German members of parliament. *Public Choice 156*(3-4), 653–676.

Persson, T. and G. E. Tabellini (2000). *Political economics : explaining economic policy*. Zeuthen lecture book series. Cambridge, Mass.: MIT Press.

Pettersson-Lidbom, P. (2008). Do Parties Matter for Economic Outcomes? A Regression-Discontinuity Approach. *Journal of the European Economic Association 6*(5), pp. 1037–1056.

Pettersson-Lidbom, P. (2012). Does the size of the legislature affect the size of government? Evidence from two natural experiments. *Journal of Public Economics 96*(3–4), 269–278.

Plott, C. R. and K. Zeiler (2005). The Willingness to Pay-Willingness to Accept Gap, the "Endowment Effect," Subject Misconceptions, and Experimental Procedures for Eliciting Valuations. *American Economic Review 95*(3), 530–545.

Plott, C. R. and K. Zeiler (2007). Exchange Asymmetries Incorrectly Interpreted as Evidence of Endowment Effect Theory and Prospect Theory? *American Economic Review 97*(4), 1449–1466.

Poterba, J. M. (1994). State Responses to Fiscal Crises: The Effects of Budgetary Institutions and Politics. *Journal of Political Economy 102*(4), 799–821.

Preuhs, R. R. (2006). The Conditional Effects of Minority Descriptive Representation: Black Legislators and Policy Influence in the American States. *Journal of Politics 68*, 585–99.

Rabin, M. (2013). An Approach to Incorporating Psychology into Economics. *American Economic Review 103*(3), 617–622.

Rao, J. M. and D. H. Reiley (2012). The Economics of Spam. *Journal of Economic Perspectives 26*(3), 87–110.

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

Rogers, J. R. (2005). The Impact of Divided Government on Legislative Production. *Public Choice 123*(1/2), 217–233.

Rogoff, K. (1990). Equilibrium Political Budget Cycles. *American Economic Review 80*(1), 21–36.

Rosenthal, A. (2009). *Engines of Democracy: Politics and Policymaking in State Legislatures.* CQ Press, Washington D.C.

Roubini, N. and J. D. Sachs (1989). Political and economic determinants of budget deficits in the industrial democracies. *European Economic Review 33*(5), 903–933.

Royed, T. J. and S. A. Borrelli (1997). Political Parties and Public Policy: Social Welfare Policy from Carter to Bush. *Polity 29*(4), 539–563.

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

Samuelson, W. and R. Zeckhauser (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty 1*, 7–59.

Schelker, M. (March 2012). Lame Ducks and Divided Government: How Voters Control the Unaccountable. *CESifo Working Paper No. 3523*.

Schram, S. F., J. Soss, and R. C. Fording (Eds.) (2003). *Race and the Politics of Welfare Reform.* Ann Arbor, MI: University of Michigan Press.

Shipan, C. R. (2006). Does Divided Government Increase the Size of the Legislative Agenda? In E. S. Adler and J. S. Lapinski (Eds.), *The Macropolitics of Congress*, Chapter 6. Princeton University Press.

Snyder, J., O. Folke, and S. Hirano (May 2012). Partisan Imbalance in Regression Discontinuity Studies Based on Electoral Thresholds. *Working Paper*.

Snyder, J. M. and D. Strömberg (2010). Press Coverage and Political Accountability. *Journal of Political Economy 118*(2), 355–408.

Soss, J., R. C. Fording, and S. F. Schram (2008). The Color of Devolution: Race, Federalism, and the Politics of Social Control. *American Journal of Political Science 52*(3), 536–553.

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.

Spiegel (2008, November 15th). http://www.spiegel.de/politik/deutschland/hamburg-niels-annen-verliert-sein-direktmandat-a-590683.html.

Strömberg, D. (2008). How the Electoral College Influences Campaigns and Policy: The Probability of Being Florida. *American Economic Review 98*(3), 769–807.

Strumpf, K. S. (2002). Does Government Decentralization Increase Policy Innovation? *Journal of Public Economic Theory 4*(2), 207–241.

Sundquist, J. L. (1988). Needed: A Political Theory for the New Era of Coalition Government in the United States. *Political Science Quarterly 103*(4), 613–635.

Svaleryd, H. and J. Vlachos (2009). Political rents in a non-corrupt democracy. *Journal of Public Economics 93*(3–4), 355–372.

Thaler, R. H. and S. Benartzi (2004). Save More Tomorrow™: Using Behavioral Economics to Increase Employee Saving. *Journal of Political Economy 112*(S1), 164–187.

The Economist (2012). *November 10th Issue*, title page.

Tsebelis, G. (1995). Decision Making in Political Systems: Veto Players in Presidentialism, Parliamentarism, Multicameralism and Multipartyism. *British Journal of Political Science 25*(3), 289–325.

Tsebelis, G. (2002). *Veto Players: How Political Institutions Work*. Princeton University Press.

Tversky, A. and D. Kahneman (1974). Judgment under Uncertainty: Heuristics and Biases. *Science 185*(4157), 1124–1131.

Tversky, A. and D. Kahneman (1991). Loss Aversion in Riskless Choice: A Reference-Dependent Model. *Quarterly Journal of Economics 106*(4), 1039–1061.

Tweedie, J. (2000). From D.C. to Little Rock: Welfare Reform at Mid-Term. *Publius 30*, 69–97.

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

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

Urban Institute (2012, April). Welfare Rules Database. Washington, D.C.

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.

US Government Spending (2013). Spending by states. http://www.usgovernmentspending.com/compare_state_spending_2010p40c. (last access: April 7th, 2014).

von Ahn, L., B. Maurer, C. McMillen, D. Abraham, and M. Blum (2008). reCAPTCHA: Human-Based Character Recognition via Web Security Measures. *Science 321*(5895), 1465–1468.

von Ahn, L. V., M. Blum, N. J. Hopper, and J. Langford (2003). CAPTCHA: using hard AI problems for security. In *Proceedings of the 22nd international conference on Theory and applications of cryptographic techniques*, EUROCRYPT'03, Berlin, Heidelberg, pp. 294–311. Springer-Verlag.

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 the Midwestern Leaders. In C. S. Weissert (Ed.), *Learning from the Leaders: Welfare Reform Politics and Policy in Five Midwestern States*, Chapter 1, Albany, NY. 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? *Princeton University mimeo*.

Wilson, T. D., C. E. Houston, K. M. Etling, and N. Brekke (1996). A new look at anchoring effects: Basic anchoring and its antecedents. *Journal of Experimental Psychology: General 125(4)*, 387–402.

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

Wiseman, M. (1996). State Strategies for Welfare Reform: The Wisconsin Story. *Journal of Policy Analysis and Management 15*(4), 515–546.

Wittman, D. (1989). Why Democracies Produce Efficient Results. *Journal of Political Economy 97*(6), 1395–1424.

Curriculum Vitae

2012-2014	Research and Teaching Assistant Collaborative Research Center “Political Economy of Reforms” and Chair of Public Finance and Economic Policy, University of Mannheim, Germany
2009-2014	PhD Student (Economics) Center for Doctoral Studies in Economics, University of Mannheim, Germany
2009-2010	Visiting Student (Economics) University of California, Berkeley, USA
2009-2011	Master of Science (Economics) University of Mannheim, Germany
2008-2009	Visiting Student (Economics) University of Copenhagen, Denmark
2006-2009	Bachelor of Science (Economics) University of Mannheim, Germany
2006	Abitur Gymnasium Neuenbürg, Germany