Accountability and the Democratic Mandate:
Analysing Pledges, Party Competition, Media Coverage, and Public Opinion

A Thesis Submitted to the Degree of
Doctor of Philosophy (Ph.D.)

Candidate
Stefan Müller

Supervisor
Professor Gail McElroy

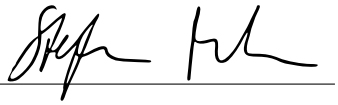
University of Dublin, Trinity College
Department of Political Science

Final Submission: March 2019

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.



Date: March 20, 2019 Stefan Müller

Abstract

The mandate model of democracy and the theory of promissory representation assume that voters sanction or reward parties for their ability to fulfil pledges made during the election campaign. While this vision of democracy has its normative appeal and although previous studies show that parties keep a large proportion of manifesto pledges, most voters do not believe that politicians fulfil their promises. I argue that the contradiction between the public perception and empirical evidence is driven by a different understanding of election pledges among citizens and scholars, and by the information voters receive and consult when evaluating the government's performance in office.

The dissertation comprises four papers. In the first paper, I reassess the established scholarly conceptualisation of election pledges using a large number of instructed crowd coders. I expect disagreement and ambiguity in pledge coding within and between the groups of experts and instructed crowd workers. Comparing the most extensive reliability exercise carried out by nine pledge scholars to over 90 crowd workers reveals that the experts have a much narrower perception of election promises. When taking the perceived vagueness of a sentence into consideration, codings between experts and the crowd become more similar. Besides raising awareness for differences in the understanding of pledges, the paper exemplifies how crowd coding can be used to reassess established concepts.

The second paper introduces the difference between 'prospective' and 'retrospective' campaign communication. Building on recent work about non-positional dimensions of party manifestos, I analyse the circumstances under which parties emphasise the past and present, and when parties instead focus on the future. A carefully validated semi-supervised classifier codes the temporal direction of over 1,200 national and subnational party manifestos. Large parts of manifestos do not address the future at all. As hypothesised, government parties employ more positive language only in statements about the past and present. Incumbents and opposition

parties are almost equally positive when it comes to describing the future. Opposition parties with low office aspiration only tend to frame the future negatively in regional elections. The study adds a new dimension to party competition and enhances our understanding of parties' campaign strategies to persuade voters with emotive rhetoric.

The third paper investigates how the media report on campaign promises. Most voters do not read manifestos or legislative text, but receive their information from news outlets. Although scholars acknowledge the importance of media for informing voters about policy outputs and outcomes, we lack evidence about the way media report on campaign pledges. The analysis of almost 500,000 statements about promises across 33 electoral cycles in four countries reveals that newspaper coverage of promises peaks before elections. News outlets indeed inform voters about parties' proposed policies. However, I observe a substantial negativity bias: newspapers report between 1.5 and 2 times more about broken than fulfilled promises. These findings help us to understand why voters often struggle to recall the fulfilment of salient pledges and why most citizens do not believe that parties keep their promises.

The fourth paper is co-authored with Tom Louwerse was published as a research note in *Political Science Research and Methods* (Müller and Louwerse 2018). We analyse how government support changes throughout the legislative cycle. Previous studies find consistent evidence for the cost of governing: incumbents lose votes at the upcoming election. However, we lack evidence as to how public support for government parties in parliamentary systems develops throughout the electoral cycle. Based on over 25,000 opinion polls from 171 electoral cycles, we show that government parties, on average, lose most support in the first half of the cycle. Second, the previously assumed curvilinear effect is more likely to occur under single-party government and in countries where the prime minister can dissolve the parliament. Third, since the 2000s, government parties rarely recover from early losses. Governing comes with high costs and that policy achievements and fulfilled promises towards the end of a legislative cycle might not increase support for governments.

Acknowledgements

First and foremost I am very grateful for the outstanding supervision and mentorship I received from Prof Gail McElroy during the past three years. Gail's responsiveness, honesty, and trust were indispensable whenever I had questions or required guidance.

I am enormously grateful for the support of two additional academic mentors and co-authors: Ken Benoit introduced me to quantitative text analysis, and offered me to join the Quanteda Initiative. I look forward to continuing my work on the development of `quanteda`-related packages and teaching infrastructure. Shaun Bowler gave extremely valuable advice on framing the dissertation chapters and thinking about additional research questions. Thanks to his enthusiasm, optimism, and great sense of humour I overcame the difficult days of writing a PhD thesis.

I also thank my other co-authors. Working with Michael Jankowski started as a coincidence and resulted in friendship and visits. I profited a lot from Tom Louwerse's insights into public opinion and data wrangling, and I am happy that we continued to collaborate after Tom moved to Leiden. The members of Quanteda Initiative, especially Kohei Watanabe, gave very valuable advice on applying methods from quantitative text analysis to campaign promises. The project with Liam Kneafsey on the impact of neutral venues on match outcomes started out of curiosity and resulted in a published paper. I will miss our coffee breaks when we analysed the weekend action in Gaelic football and hurling.

My dream of living in Dublin and writing a PhD thesis at Trinity College would not have been possible without financial support from various sources. First, I thank the Studienstiftung des deutschen Volkes (German Academic Scholarship Foundation) for supporting me throughout my undergraduate studies and for funding my master's degree at Trinity College Dublin. Second, I thank Trinity College for selecting me for a Postgraduate Ussher Fellowship. Third, I am grateful to the Irish Research Council for funding this dissertation project with a Government of Ireland Postgraduate Scholarship.

Before, during, and after conferences and workshops I received very helpful feedback from Ken Benoit, Shaun Bowler, Royce Carroll, Patrick Chester, Alan Duggan, Jason Eichorst, Zachary Green, Martin Gross, Alexander Horn, Robert Huber, Heike Klüver, Liam Kneafsey, Nicolas Merz, Christian Rauh, Roman Senninger, Shane P. Singh, Miriam Sorace, and Mariken van der Velden. The members of the Comparative Party Pledges Group – especially Rory Costello, Terry Royed, and Robert Thomson – were interested in my methodological approach, sent me the expert coding results, and offered very valuable insights from the perspective of pledge scholars. Without their support, parts of this thesis could not have been written.

The Friday seminar at Trinity College Dublin provided great opportunities to present drafts of my work. I thank the staff members and PhD students for very helpful comments on early versions of my dissertation papers. Thanks to my PhD cohort – Alan, Ciarán, Silvia, Somya, and Raluca – for many conversations about our research and especially for the chats about topics that were not related to our PhDs.

I would like to thank Fabrizio Gilardi for offering me a postdoctoral research position before I finished my PhD thesis. I am excited to start working at the University of Zurich with such supportive and productive scholars.

It is great that my friendship with Alex, Paddy, and Roman, my best friends from the master's course in 2014/15, is still so strong. I hope we will continue to visit each other and travel to Ireland together, although we now live all over Europe. I thank my Irish housemates, especially John and Fergus, for dozens of chats, a lot of fun, and many movie nights in our living room. I will never forget the Saturday afternoons and evenings watching German Bundesliga and discussing the latest transfer rumours with Christian, Dennis, Leo, Lion, and Tobi. You all made my stay in Ireland an unforgettable one. A big thank you goes to my friends and relatives in Bonn and Cologne, and the members of the TSV Bonn rrh. who always welcomed me with open arms. Echte Fründe ston zesamme!

I dedicate this PhD thesis to my parents Rosi and Bernd, and my sister Simone. You helped me so much throughout my entire life, and made sure that I maintain a healthy work-life balance. Thank you for everything!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Modern democracies rest on the indirect representation of voters' preferences through elected candidates. Parties propose policies during election campaigns, and voters choose among these alternatives when casting their ballot. Elections empower voters to hold politicians accountable for past achievements or failures, or to vote prospectively based on the promises for an upcoming legislative cycle. Through regular elections, voters authorise representatives to implement policy (Downs 1957; Klingemann et al. 1994; Powell 2000; Mansbridge 2003; McDonald and Budge 2005). Once in office, representatives should aim to fulfil their pre-electoral promises.

Political parties take a crucial role in enabling a relationship between voters and representatives. Parties select candidates and organise elections, mobilise voters, shape the political discourse, offer alternative policies, and finally try to deliver the promises made prior to an election (Dalton et al. 2011). This programme-to-policy linkage lies at the heart of the 'mandate model of democracy'. The mandate model requires that parties offer different policy proposals, that voters know about fulfilled and broken pledges and parties' promises for the future, and that parties aim to fulfil their pledges (American Political Science Association 1950; Manin et al.

1999; McDonald and Budge 2005; Roberts 2010). This theory of democracy enjoys popularity as a desirable mode of representation. As Manin et al. (1999: 30) state, "the mandate conception of representation is widespread: scholars, journalists, and ordinary citizens rely on it as if it were axiomatic".

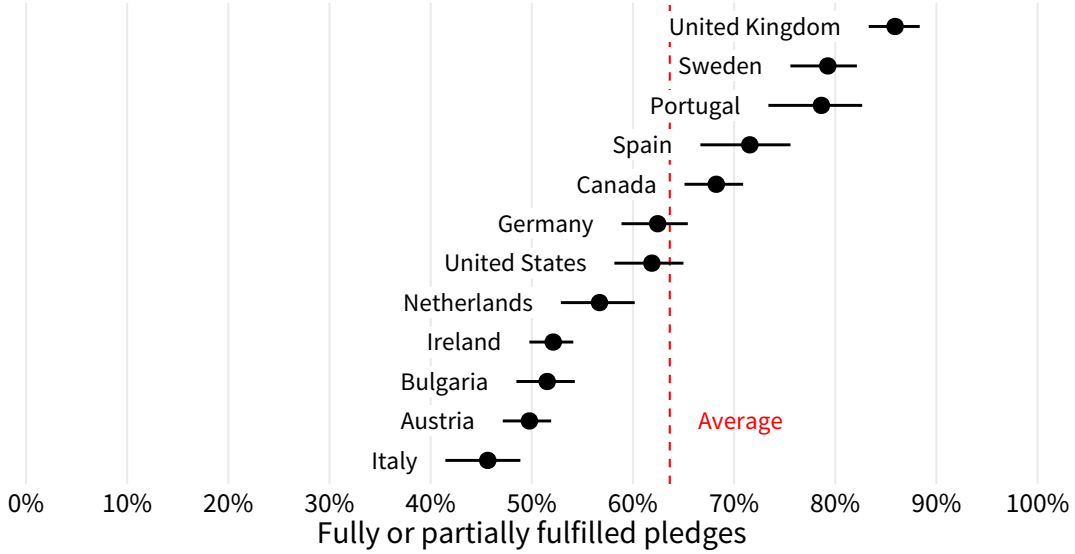
But do parties fulfil their promises, and do voters know about fulfilled and broken pledges? These questions are of pressing importance, as political parties "seem to be experiencing potentially severe legitimacy problems and to be suffering from a quite massive withdrawal of popular support and affection" (Mair 2008: 211). Declining levels of turnout, rising electoral volatility, and decreasing levels of party identification and party membership underscore that the linkage between voters and parties has come under pressure (Dalton and Wattenberg 2000; Mair 2013). Feelings towards politicians, parliament, and councils have also become substantially more negative in recent decades (Clarke et al. 2018). Citizens increasingly lack trust in politicians and their ability to keep their promises.

Despite public dissatisfaction with parties and politicians, election pledges are a still a central topic in news coverage and political debates. Not fulfilling important pledges or making unpopular promises can result in electoral defeats. For instance, the US presidential candidates' promise to expand health care in the form of 'Obamacare' and the subsequent pledge of repealing this policy received massive media attention and polarised the electorate during three presidential campaigns. In the 2015 general election, the Liberal Democrats in the United Kingdom suffered heavily from breaking the promise of not increasing tuition fees for third-level education. Prior to the 2013 general election, the German Green party completely underestimated public outrage when promising to introduce a weekly 'veggie day' in canteens. These cases exemplify that parties have to consider carefully what to promise and how to implement these pledges in government.

Many studies have attempted to measure mandate fulfilment across countries, over time, and conditional on institutional settings. The results are quite encouraging.

Parties fulfil a large proportion of pledges specified in election manifestos (Pomper and Lederman 1980; Rose 1984; Kalogeropoulou 1989; Royed 1996; Thomson 2001; Naurin 2011; Artés and Bustos 2008; Artés 2013; Moury and Fernandes 2018; Praprotnik 2017). Figure 1.1 plots the percentage of fully or partially fulfilled pledges from the first large-scale comparative study on pledge fulfilment (Thomson et al. 2017). Across 57 campaign in 12 countries, parties fulfil around 60 per cent of their manifesto pledges, with this figure exceeding 75 per cent for the United Kingdom, Sweden, and Portugal.

Figure 1.1: Percentage fulfilled or partially fulfilled pledges by government parties

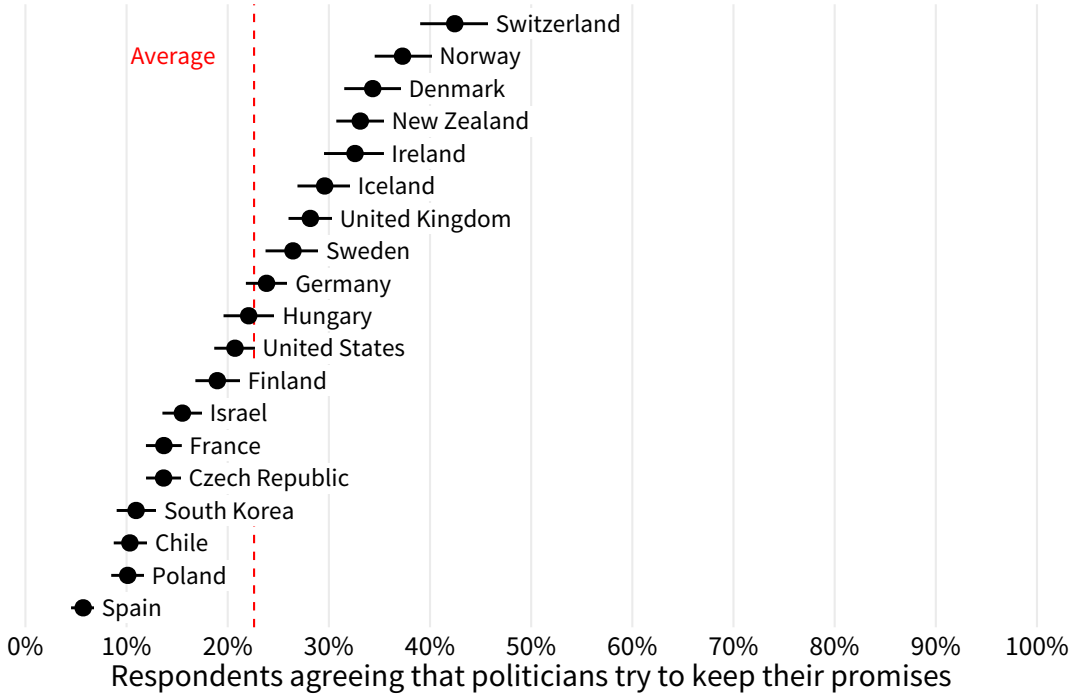


Note: Figure based on data by Thomson et al. (2017). Error bars show 95 per cent bootstrapped confidence intervals.

However, puzzlingly, most citizens do not believe that parties keep their pledges (Naurin 2011; ISSP Research Group 2018). In the most recent cross-national survey, on average only 22 per cent of respondents agree with the statement "People we elect as MPs try to keep the promises they have made during the election" (Figure 1.2). Moreover, many voters struggle to recall the fulfilment of salient campaign pledges. Across five studies asking citizens about the fulfilment of 29 promises, on average 57 per cent of the respondents made an incorrect evaluation or did not know whether a

specific pledge has been fulfilled (Thomson 2011; Thomson and Brandenburg 2018; Naurin and Oscarsson 2017; Belchior 2018; Duval and Pétry 2018a). The inaccurate perception of pledges has serious consequences for the validity of the mandate vision of democracy. While scholars find a high degree of mandate fulfilment, citizens do not believe in pledge fulfilment and struggle to recall whether a handful of central promises have been kept or broken.

Figure 1.2: Percentage of respondents believing that politicians keep their promises



Note: Figure based on data from ISSP Research Group (2018). The plot displays the unweighted proportions of respondent who agree or strongly agree to the statement "People we elect as MPs try to keep the promises they have made during the election". 'Don't know' responses are excluded. Error bars show 95 per cent bootstrapped confidence intervals.

The contradiction between public scepticism and the academic findings regarding pledge fulfilment has been termed the 'pledge puzzle' (Naurin 2011). This dissertation aims to better understand this puzzle by offering a novel perspective on political accountability and the democratic mandate. More precisely, I tackle four research questions: *First*, do citizens and experts perceive the concept of election pledges

differently? *Second*, under what circumstances do parties focus on describing future policy and when do parties engage in credit claiming and blame attribution? *Third*, could media coverage of election promises contribute to the common public view of 'promise-breaking politicians'? *Finally*, at what times during the electoral cycle do voters hold governments accountable in opinion polls, and what institutional factors condition this effect? Finding answers to these questions sheds light on the question of whether or not promissory representation works in practice and why citizens have become so sceptical regarding parties' willingness and ability to implement their pre-electoral pledges.

1.1 Research Gaps and Contributions

In the following subsections, I summarise previous evidence and point to the four research gaps that the chapters of this dissertation address.

1.1.1 The Validity of Coding Campaign Pledges (Chapter 2)

Scholars have made major progress in deriving a common definition of pledges in order to compare results from single-country studies (Naurin 2014; Thomson et al. 2017; Royed et al. 2019). The Comparative Party Pledges Group (CPPG), a collaboration between leading pledge researchers, has conducted reliability exercises to test whether scholars code a set of manifesto statements in the same way. Despite acceptable pairwise agreements, many statements are not coded in the same way by all scholars. The classification of manifesto statements as pledges is difficult and ambiguous. On purpose, parties often remain vague when drafting manifestos and do not explicitly state how a future action will be fulfilled (Rovny 2012; Horn and Jensen 2017; Eichorst and Lin 2019). Additionally, the question of the concept validity of election pledges has been largely overlooked. We need to know whether scholars perceive pledges differently than non-expert coders, in order to interpret the extent of

fulfilled promises (Dupont et al. 2019). Using recent advancements in crowd-sourced text coding (Benoit et al. 2016; Haselmayer and Jenny 2017; Horn 2018; Lehmann and Zobel 2018), in this chapter I revisit the concept of pledges. To test whether non-experts code manifesto statements in a similar way to experts, workers receive detailed instructions and code the same set of sentences as the pledge scholars. The results underline that scholars apply a much narrower definition of pledges which may explain why the proportion of fulfilled promises reported in academic studies is higher than assumed by the public.

1.1.2 Retrospective and Prospective Campaign Communication (Chapter 3)

Voters attribute blame and credit to parties (Weaver 1986; Marsh and Tilley 2010; Burlacu et al. 2018). As a consequence, before elections, parties should not only focus on the future, but also make assessments of the present and past. Surprisingly, however, we lack knowledge about parties' strategic behaviour in focusing on retrospective actions or failures, and proposals for future action. Chapter 3 analyses how parties engage in credit claiming, blame attribution, and the framing of the future in official campaign communication.

Despite the extensive usage of party manifestos to study left-right positions, issue ownership, or election pledges, "the processes of manifesto production, enactment, and public reception are not very well understood" (Dolezal et al. 2012: 869). Only recently, scholars have started to observe non-positional aspects of campaign communication (Somer-Topcu 2015; Lo et al. 2016; Bräuninger and Giger 2018; Crabtree et al. 2018; Eichorst and Lin 2019; Harmel 2018). Previous research does not distinguish between statements about the future and sentences about the past and present (for an exception see Dolezal et al. 2018). Instead, some studies explicitly equate manifestos with a collection of statements about the future (Horn and Jensen

2017; Alexiadou and Hoepfner 2018; Kosmidis et al. 2018). Yet, we do not know what proportions of manifestos actually address future policies.

In this chapter, I predict that parties place considerable emphasis on describing the present and past in their manifestos. Because voters obviously struggle to recall the fulfilment or breaking of promises, parties face incentives to add statements that claim credit or attribute blame, hoping that journalists report on their own achievements or their opponents' failures. Classifying over 1,200 national and subnational party manifestos, Chapter 3 introduces prospective and retrospective rhetoric as an important dimension of campaign communication. On average, only one in two manifesto sentences relates to the future. Incumbency status strongly influences credit claiming and blame attribution. Finally, opposition parties on the *national* level rarely describe the future negatively. However, parties with low office aspiration on the *regional* level are more likely to develop 'doomsday' campaigns.

1.1.3 Promissory Representation and the Media (Chapter 4)

The media play a crucial role in the mandate vision of democracy. News outlets provide a link between parties and voters, and serve as agenda setters and gatekeepers. Moreover, the media inform citizens about parties' (in-)ability to fulfil specific pledges (Håkansson and Naurin 2016; Prior 2017; Soroka and Wlezien 2018). Journalists pay close attention to policy outputs and inform citizens about policy outputs and outcomes (Arnold 2004; Prior 2017). For instance, Håkansson and Naurin (2016: 395) state that "[f]or the news media, election promises form an ideal basis of scrutiny for the fulfillment of their watchdog ideal". Election promises provide a suitable way of covering political events, and engaging in horse race or 'game frame' journalism. Since the "average citizen often does not experience government policy directly, but learns about it from the mass media" (Soroka and Wlezien 2018: 1), media coverage of promises can strongly influence public opinion.

Although scholars seem to recognise the role of the media when assessing promissory representation, only one study to date, limited to the four weeks prior to Bulgarian elections, assesses which promises are reported in newspapers (Kostadinova 2017).[1] We lack knowledge as to whether coverage of promises follow certain patterns across countries and throughout the cycle, and whether newspapers exert a negativity bias when reporting on promises. Chapter 4 provides the first comparative study on media coverage of political promises. Classifying over 480,000 statements on promises from newspaper articles across 33 electoral cycles in Australia, Canada, the Republic of Ireland, and the United Kingdom reveals that parties cover promises throughout the entire cycle. However, news outlets include more articles on broken than fulfilled promises. Whereas voters usually do not read manifestos and struggle to recall the fulfilment of salient promises, journalists rarely cover the successful fulfilment of promises.

1.1.4 Government Support Between Elections (Chapter 5)

The fourth research gap identified in this thesis relates to the development of support for government parties throughout the electoral cycle. Downs (1957: 41) already noted that "every election is a judgment passed upon the record of the incumbent party". According to Powell (2000: 11), the power of the voters creates "a pressure on all incumbents to worry about the next elections and make policy with voter review in mind". While a large body of work analyses responsiveness during the electoral cycle (Stimson et al. 1995; Wlezien 1995; Soroka and Wlezien 2010), we lack systematic evidence about changes in approval for government parties throughout the legislative cycle.

Many studies clearly show that incumbents lose voters in the upcoming election (Nannestad and Paldam 1999; Stevenson 2002; Wlezien 2017). Cuzán (2015) even

[1] Kostadinova's (2017) single-country study also only takes into consideration the month before an election, which does not allow for an analysis of coverage of promises throughout the cycle.

describes this phenomenon as a 'law of politics'. What we do not know, is the time throughout a cycle when governments lose most support and whether incumbents recover prior to an upcoming general election. Previous studies on the United States find that the party of the president suffers in midterm elections, but prior to *Election t+1*, the president tends to regain support, which reduces the net losses (Mueller 1970; Tufte 1975; Erikson 1988; Campbell 1991; Folke and Snyder 2012). Similar evidence of this curvilinear 'electoral cycle effect' was found in European Parliament elections (Reif and Schmitt 1980) and German state elections when the upper and lower house are controlled by the same coalition (Jeffery and Hough 2001; Kern and Hainmueller 2006). However, determining national-level government support through second-order election results is problematic. In many countries, only a few relevant elections take place between two first-order elections. Moreover, national and sub-national elections are not determined by the same policies, and are held under different, regionally specific cleavages.

Chapter 5 maps party support throughout the electoral cycle by analysing parties' standing in national opinion polls. Compared to second-order elections, opinion polls offer snapshots of public support on a more regular basis.[2] Analysing opinion polls from 171 electoral cycles, I find that 'clarity of responsibility' (Powell and Whitten 1993) conditions the electoral cycle effect. When a voter knows who is to blame for a policy, it is more likely that a curvilinear effect in support occurs. Moreover, higher dissolution powers of the prime ministerial party correlate with a larger recovery at the end of an electoral cycle. Leveraging polling data for several countries since the 1960s, the chapter shows that parties increasingly struggle to regain support since the 2000s.

[2]Two recent studies also analyse government support over time, but either do not focus on institutional or election-specific circumstances (Green and Jennings 2017) or mainly analyse the effect of party support in polls on the calling of early elections (Walther and Hellström 2019).

1.2 Relevance and Implications

To summarise, this thesis explores the mandate model of representation from a series of new perspectives. Previous research clearly shows that some of the main assumptions of the mandate model of democracy do not hold in practice: many voters are not aware which promises a government has broken or fulfilled; and only few voters have an accurate understanding of the promises made in party manifestos, the major source of previous studies. Shifting the methodological focus from manual content analysis to crowd coding and quantitative text analysis of large text corpora allows us to improve our understanding of promissory representation. On the one hand, pledges are often vague. Specific and 'objectively' testable pledges, included in academic studies, on the other hand, might not receive much public attention. The thesis points to the need to move beyond the analysis of manifesto content. Pledges covered in the media and made by party leaders in televised debates require much closer attention, since these statements are more likely to influence vote choice. When coding promises, researchers should aim to collect multiple codings per sentence, and add measures of uncertainty as well as information on the saliency of promises.

The analysis of opinion polls demonstrates that governments struggle to recover from early losses in the electoral cycle. For this reason, parties may increasingly point to their own achievements prior to general elections (see, e.g., Fianna Fáil 2002; SPD Bundestagsfraktion 2017). We need to analyse in greater detail when parties decide to claim credit or allocate blame instead of making pledges for the future. More attention should also be devoted to the effects of retrospective rhetoric on citizens' perceptions and voting behaviour. In times of rising dissatisfaction with established parties and politicians, this thesis makes important contributions to the public and scholarly discussion about the linkages between voters, parties, and the media.

Chapter 2

Reassessing an Established Concept Through Crowd-Sourced Text Coding

Abstract

A growing body of research analyses whether crowd workers can reproduce the 'gold standard' of expert-generated data. Based on the case of election pledges, I show how crowd-coding can also test for differences in perceptions of a concept between groups of experts and instructed non-experts. Comparing the most extensive reliability exercise carried out by nine pledge scholars to 3,660 codings generated by 90 crowd workers reveals considerable disagreement within and between both groups. Moreover, the carefully instructed and continuously monitored non-experts have a much broader understanding of election promises, which has important implications for analysing pledge fulfilment. The approach illustrates that crowd-coding could be used across all subfields of political science to reassess the measurement validity of a concept.

2.1 Introduction

The formation of concepts lies at the heart of social science research (Adcock and Collier 2001). A research design lacks validity without a reliable operationalisation of the concept under investigation (Krippendorff 2004; Gerring 2012). In political science, many concepts are measured based on human interpretation of text. For instance, in the Comparative Manifesto Project (CMP/MARPOR) human coders classify policy issues in party manifestos, and democracy indicators such as Polity IV rely on human judgement. Although multiple codings are required to determine reliability and uncertainty (Mikhaylov et al. 2012), for reasons of expense, convenience, or recruitment researchers often have only a single coding per unit. Therefore, political scientists have recently turned to crowd coding in order to circumvent the problem of too few experts. In crowd-sourced coding tasks, the observations are divided up into small units of work. These groups of observations are then assessed and scored by large numbers of paid and instructed workers. Amongst other things, crowd workers have been employed to classify political events, sentiment, news, or party manifestos (e.g., D'Orazio et al. 2016; Benoit et al. 2016; Haselmayer and Jenny 2017; Lind et al. 2017; Horn 2018; Lehmann and Zobel 2018). In almost all of these studies, the performance of crowd coders is assessed against the 'gold standard' of expert coding. The scholarly conceptualisation is used as the benchmark for the quality of the crowd. However, the comparison of multiple crowd and expert codings can also uncover problems with a prevailing measurement of a concept. In other words, the method reveals whether experts perceive a concept in a fundamentally different way than non-expert coders. Researchers can use this evidence to refine the measurement validity of a concept.

This paper makes two contributions to the literature. First, I test for differences in the measurement of election pledges by experts and instructed non-experts. In the most comprehensive reliability exercise, nine leading pledge scholars have coded a set

of 138 statements from party manifestos. All participants in the coding exercise are authors of the recent comparative study on pledge fulfilment (Thomson et al. 2017).[1] To further explore the reliability, I collected multiple crowd codings for the same set of statements. Comparing experts' pledge assessment to 3,660 crowd-codings reveals that non-experts have a much broader perception of election promises than experts, even though the crowd received the same instructions, had to pass an entry test, and was subject to continuous attention checks. The results remain very stable once more than four crowd codings per sentence are collected. The broader understanding by the crowd has wide-ranging implications for the empirical assessment of promissory representation (Mansbridge 2003). The observation that experts apply a narrower definition of pledges, could help to explain why the public does not believe that parties keep their promises. (ISSP Research Group 2018). Whereas scholars only perceive objectively testable statements as pledges, voters might also perceive statements about the future with lower specificity and without a binding force as pledges. A survey experiment reveals that respondents are more likely to perceive messages with a binding force as pledges. Against the expectation, vague statements were *more* likely to be perceived as pledges than very specific statements (Dupont et al. 2019).[2] Scholars, however, do not consider these statements in their analysis. As a consequence, existing studies may employ a concept that does not correspond to the public perception of the pledges. Dupont et al.'s (2019) findings support the main conclusion of this paper: vague statements with a reference to the future might be regarded as pledges by non-experts, but not by pledge scholars.

To be clear, crowd workers are not representative of any population of voters or citizens. Therefore, the findings of this paper cannot be interpreted as thee *public*

[1] I am very grateful to the Comparative Party Pledges Group for providing the expert coding decisions. Personal information about the scholars cannot be retrieved from the data, since I only have aggregated counts on the sentence-level.

[2] Although statistically significant, the treatment effects of 0.23 (for specificity) and –0.27 (for binding force) can be considered small, amounting to less than 25 per cent of the standard deviation (1.04).

perception of pledges. Yet, the findings reveal that scholars define the concept of pledges much more narrowly. Future studies could rerun a similar coding experiment with representative samples and check whether the results persist when asking voters to code manifesto statements.

Second, the paper contributes to the body of work on the systematic analysis and measurement of concepts. Based on the example of election pledges, I show how crowd-sourced text analysis can be used to understand whether scholars and instructed non-experts assess a concept differently. Referring to Adcock and Collier's (2001) levels of measurement, especially the 'background concept' (the broad understandings of a given concept) and the 'systematised concept' (a specific formulation of the concept) can be reassessed with the method introduced in this paper. Moreover, researchers can investigate how different coding instructions or definitions influence the scoring of a concept (Crepaz and Chari 2018). If the meaning of a concept is disputed among scholars and instructed non-experts, researchers may revisit broader issues concerning the background concept. The study illustrates that crowd-coding could be used across all subfields of political science as a fast and affordable tool for developing or reassessing concepts.

2.2 Conceptualising Election Pledges

Scholars define an election pledge as "a statement committing a party to one specific action or outcome that can be clearly determined to have occurred or not" (Thomson et al. 2017: 532). Importantly, a promise must be testable in order to count as a pledge. But do coders without knowledge of the academic literature have a similar perception of promises when coding manifesto sentences based on the scholarly definition? The answer to this question speaks to the so called 'pledge puzzle' (Naurin 2011): while previous studies find shares of fulfilled promises ranging between 40 and 85 per cent (e.g., Royed 1996; Thomson 2001; Praprotnik 2017; Thomson et al. 2017), a recent

representative survey across 28 democracies reveals that only around one in five respondents believes that politicians try to keep their promises (ISSP Research Group 2018). This contradiction might evolve from a fundamentally different understanding of election promises. Therefore, a critical assessment of the conceptualisation of pledge fulfilment seems necessary, and election pledges constitute a suitable case to apply the crowd-sourced concept reassessment.

2.2.1 Expert Coding of Pledges

Researchers working on election promises have made considerable progress in harmonising pledge coding and deriving a common definition and coding scheme. The Comparative Party Pledges Group (CPPG), an international consortium of pledge researchers, aims to improve comparability across single-country studies. When the scholars started to focus on a comparative approach, they developed a common definition of election pledges, outlined detailed coding instructions, and presented examples for statements that (do not) fulfil the requirements of pledges (see extensively the Supporting Information of Thomson et al. 2017). For cross-country studies that involve data coded by many authors (Thomson et al. 2017; Naurin et al. 2019) the group has set up coding reliability tests. In the group's most extensive coding exercise 12 pages of the 2008 manifesto by the Conservative Party of Canada were coded by nine scholars (see Section A.3).[3] I have information about how many scholars coded the respective statement as a pledge for each of the 138 sentences comprising these 12 pages. These recent efforts in studying pledge fulfilment from a comparative perspective allow for a comparison of the scholarly conceptualization to the coding by non-experts. Since almost all leading pledge researchers have participated in this reliability test, the codings can be regarded as a representative sample of the scholarly understanding of election pledges.

[3]As a point of comparison, the reliability exercise of CMP codings by Mikhaylov et al. (2012) involved 179 quasi-sentences.

2.2.2 Crowd Coding of Pledges

Crowd-sourcing offers a reproducible, transparent, fast, and cost-effective way of coding political text (Benoit et al. 2016). Online workers have gained more weight in social science in general, especially for experimental designs (e.g., Berinsky et al. 2012; Mullinix et al. 2015; Huff and Kertzer 2018) and the annotation of political text (Benoit et al. 2016; Horn 2018; Lehmann and Zobel 2018). Using crowd coders (instead of, for instance, collaborators or research assistants) for reassessing a political science concept has three advantages. First, an entry test and continuous quality checks ensure that workers do not code statements according to their personal views, but stick to the detailed instructions. Coders who fall below a threshold of correctly answered question (see below) are automatically removed from the job (Berinsky et al. 2014). Second, reliability should not depend on prior experience. Instead, "agreements or disagreements among multiple descriptions [must be] generated by a coding procedure, regardless of who enacts that procedure" (Krippendorff 2004: 414). Third, coders with domain-specific knowledge tend to be influenced by cues when evaluating ambiguous statements, which might induce bias (Ennser-Jedenastik and Meyer 2018). Crowd workers without specific knowledge of a case or country are less likely to be influenced by these cues. Thus, crowd-sourcing constitutes an excellent environment for reassessing an existing concept.

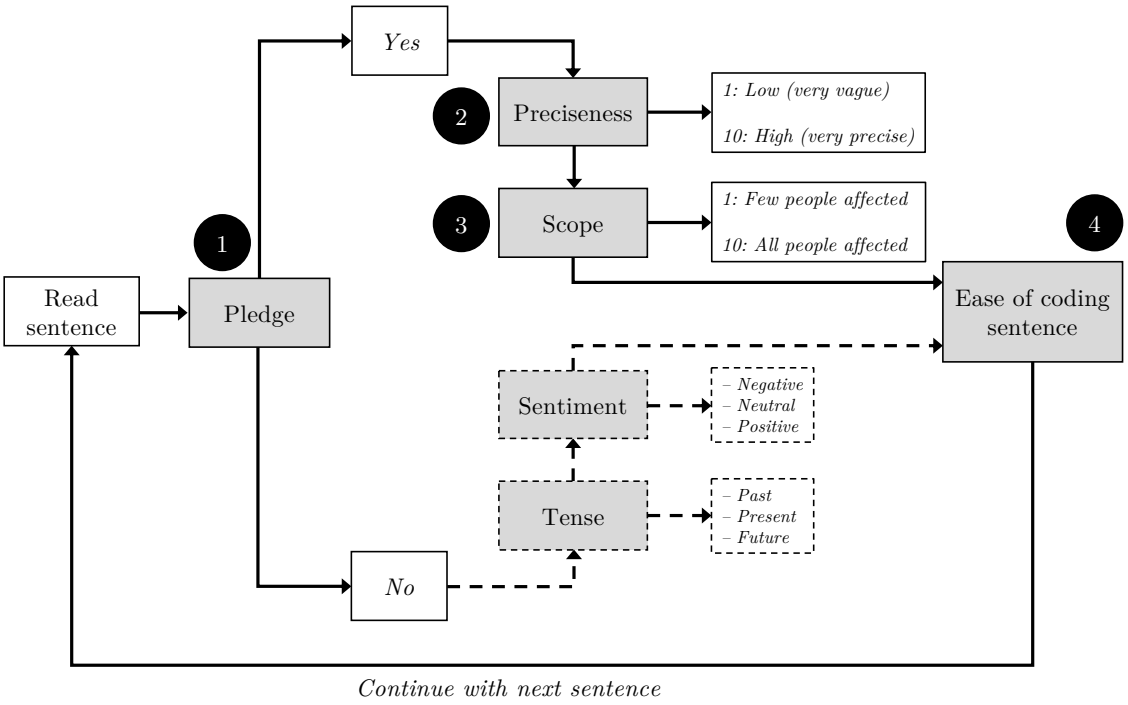
Workers are recruited, and data are collected on the online platform CrowdFlower (in 2018 renamed to Figure Eight).[4] Potential contributors first read detailed coding instructions with several examples (Section B.2). Workers are required to code strictly according to the instructions, not based on their personal definition of an election pledge. Afterwards, workers need to answer four out of five questions correctly.

[4]https://figure-eight.com. The platform has more than 10,000 registered workers from all over the world, and has been used extensively by political scientists (e.g., Benoit et al. 2016; Horn 2018; Lehmann and Zobel 2018).

Having passed this quiz, respondents code randomly selected sentences from the text corpus in groups of five questions.

The coding process works as follows (Figure 2.1): contributors read a sentence and have to decide whether the statement should be coded as a pledge according to the definition provided in the instructions (Step 1 in Figure 2.1). If a statement is classified as a pledge, the worker needs to determine the preciseness (Step 2) and scope of the pledge (Step 3). The preciseness of the election promise ranges from 1 (very vague) to 10 (very precise). The scope depends on how many people or groups of society will be affected by a pledge, and ranges from 1 (few people affected) to 10 (almost all people affected). Sentences not coded as a pledge are classified in terms of the prevailing tense and the sentiment. In a last step, coders are asked how easy or difficult they perceived the coding of the statement.

Figure 2.1: Flowchart of the CrowdFlower coding procedure



Crowd-sourced text analysis requires test questions (also called gold tasks) as a control system to remove 'spammers' (Berinsky et al. 2014; Benoit et al. 2016). For these test questions, the 'answer key' (how the statement should be coded

correctly) is specified in advance. 20 per cent of all sentences are test questions occurring at a random position in each block of five statements to be evaluated. I select test questions from an expert reliability test of the 2002 manifesto of the Irish party Fianna Fáil that are coded identically by all expert coders and clearly do (not) fulfil the basic requirements of a pledge.[5] Following the recommendations by Berinsky et al. (2014) and Benoit et al. (2016) I also add 11 'screeners' to the set of test questions. Screeners are sentences with exact instruction on how to code a statement (surrounded by two actual manifesto sentences) to ensure that respondents pay attention.[6] Workers need to answer 80 per cent of the test/screener questions correctly throughout the job – a benchmark based on existing studies (Benoit et al. 2016; Lind et al. 2017). CrowdFlower participants are limited to English speaking countries, needed to have completed at least 100 test questions in earlier jobs, and have had an overall rate of 80 per cent correctly answered test questions throughout their prior coding jobs. These restrictions are supposed to ensure a certain standard of familiarity with English and to guarantee that coders know the usual procedure of crowdcoding jobs.

I selected a minimum of 20 crowd judgements for each of the 138 sentences (Benoit et al. (2016) recommend a minimum of 5 codings per sentence). I aggregate the codings for the crowd and experts by estimating the proportion of coders from each group that coded a statement as a pledge. For the continuous scales of preciseness and scope, I calculate the average scores, which can theoretically range between 1 and 10. Since crowd workers are paid in an interval of five coding decisions, the number of codings per worker differs substantially. To limit the impact of individual coders and to circumvent the problem of fatigue, I set a limit of 80 codings per worker (see for a similar approach Horn 2018).

[5] I followed the guidelines established by CrowdFlower: https://success.figure-eight.com/hc/en-us/articles/202702985-How-to-Create-Test-Questions.

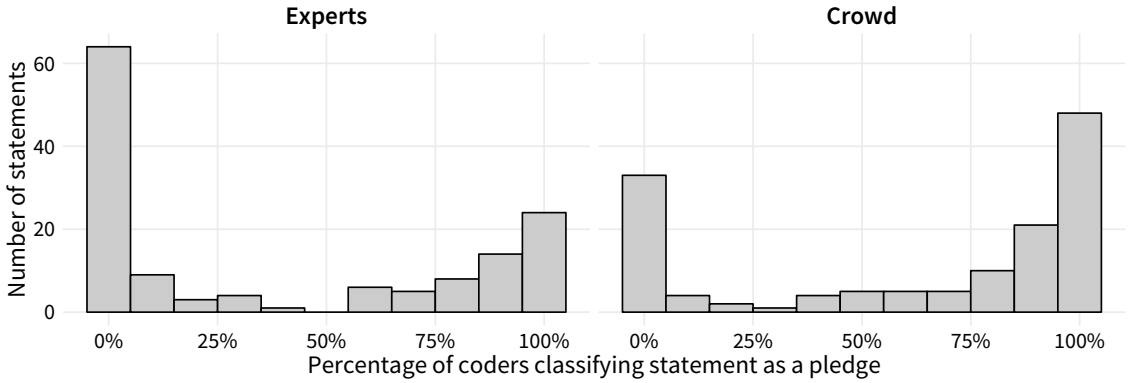
[6] For example: "Code this sentence as a Pledge with a Preciseness of 7, a Scope of 6 and Easy."

2.3 Results

The presentation of the findings proceeds in three steps. First, I compare disagreement within the groups of scholars and crowd workers. Second, I assess (dis)agreement between the aggregated codings of both groups. Third, using a bootstrapping approach I highlight that the conclusions depend on the number of codings per sentence and whether the perceived vagueness of a statement changes the aggregated assessment by the crowd.

Figure 2.2 compares the distribution of the pledge classifications for experts and the crowd. A score of 0 per cent (100 per cent) indicates that all coders of the group assigned the label 'no pledge' ('pledge') to a sentence. When half of the coders code a statement as a pledge, while the other half does not regard the sentence as a pledge, the evaluation equals 50 per cent (the highest possible value of internal disagreement). Section A.3 lists all sentences along with the aggregated average pledge evaluations by scholars and crowd workers. The histograms follow a bimodal distribution for experts and crowd workers indicating that many statements were coded as pledge/no pledge by all coders of each group. Importantly, crowd workers coded almost twice as many sentences as a pledge when compared with the nine scholars.

Figure 2.2: Comparing the distributions of aggregated pledge evaluations by experts and crowd workers



Yet, we also observe substantial disagreement within each group.[7] 50 out of 138 statements (36 per cent) were not coded in the same way by all scholars. The disagreement reduces to 27 statements (19 per cent) when we only consider statements as disputed that were coded differently by between 20 and 80 per cent of the experts. For crowd workers we observe very similar patterns. 50 per cent of the statements were not coded into the same category (pledge/no pledge) by all workers when considering 20 crowd codings per sentence.

Having outlined the distribution of codings within each group, Figure 2.3 plots the aggregated evaluations by the experts and crowd for all statements. 35 per cent of the statements (the points at 0%/0% and 100%/100%) were coded in the exact same way by all experts and crowd workers. However, 48 per cent of the statements were coded as a pledge by a substantially larger proportion of crowd workers than experts (the aggregated differences exceed 0.2). For only 3 statements, the aggregated expert evaluation exceeds the aggregated crowd evaluation. 29 statements (21 per cent) are 'disputed' within both groups, i.e., at least one expert and one crowd worker deviates from the coding by the majority. For instance, the following statement was coded by 33 per cent of the experts and 65 per cent of the crowd workers as a pledge, and exemplifies the difficulty of pledge coding: "New competition laws will make it easier to investigate and prosecute bid-rigging and hard-core cartel behaviour such as price fixing." The 'disputed' statements are listed in Table 2.1. Without multiple codings per unit, it is not be possible to detect these obviously ambiguous cases.

Table 2.1: Disputed statements among experts and crowd coders

Statement	Pledge mean (experts)	Pledge mean (crowd)
Reaffirming the Ban on the Bulk Transfer of Water We will reaffirm the Canadian Government's position that NAFTA cannot require Canada to export bulk water to other NAFTA countries.	0.56	0.92
Promoting Biofuels, Wind and Other Alternatives The Conservative Government is investing $1.5 billion over the next seven years in the production of biofuels, and requiring gasoline to contain 5 per cent renewable content by 2010 (2 per cent renewable content for diesel by 2012).	0.33	0.78

[7]Note that scholars sometimes detected more than one pledge per sentence. On the aggregated level, I selected the most frequently detected pledge per sentence.

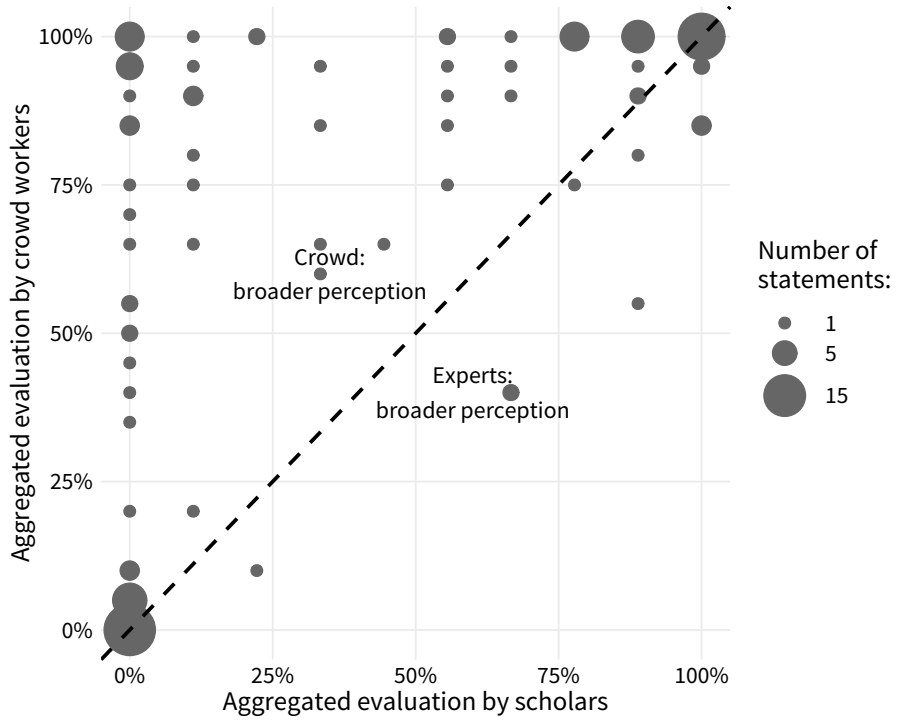
Table 2.1: Disputed statements among experts and crowd coders

Statement	Pledge mean (experts)	Pledge mean (crowd)
We are investing $1.48 billion over four years into incentives to produce more wind, solar, geothermal and tidal power - an investment that will produce over 14 million megawatt hours of clean electricity, or enough to power a million homes.	0.11	0.75
Generating 90% of Our Electricity From Non-Emitting Sources by 2020 As part of our Turning the Corner action plan, a re-elected Conservative Government will work to ensure that 90 per cent of Canadian electricity needs are provided by non-emitting sources such as hydro, nuclear, clean coal or wind power by 2020.	0.56	0.95
The Conservatives and Stephen Harper believe that the current Senate must be either reformed or abolished.	0.22	0.08
We will work to ensure that appointees to federal agencies, boards, commissions and Crown corporations reflect the diversity of Canada in language, gender, region, age and ethnicity.	0.33	0.94
The Commission will oversee the selection process for appointments to federal agencies, boards, commissions and Crown corporations, as well as develop guidelines, review and approve the selection processes for appointments and report publicly on the Government's compliance with the guidelines.	0.11	0.85
We will respect the jurisdiction of the provinces and territories in the Constitution Act, 1867, and will enshrine our principles of federalism in a new Charter of Open Federalism.	0.11	0.91
At least 25 per cent of the CRTC commissioners will be French-language speakers, appointed in consultation with Quebec and groups representing linguistic minorities.	0.89	0.57
Hearings related to Frenchlanguage or Quebec broadcasters will be heard by panels consisting of a majority of French-language or Quebec CRTC members.	0.67	0.37
We will take steps to help celebrate Canadian history and identity and develop a stronger sense of national citizenship.	0.11	0.95
We will continue support for sport and amateur fitness in Canada, at both the elite and recreational levels.	0.11	0.85
New competition laws will: Make it easier to investigate and prosecute bid-rigging and hard-core cartel behaviour such as price fixing.	0.33	0.70
Raise maximum penalties for bid-rigging and cartels to a $25-million fine and 14 years in prison.	0.89	0.93
Introduce fines of up to $10 million - $15 million for repeat offenders - for companies that abuse their dominant market position.	0.89	0.89
Provide for restitution for consumers who fall victim to deceptive marketing practices.	0.56	0.78
The new law will reduce dangerous, destructive and deceptive email and web site practices, and will establish new fines for those who break the law.	0.44	0.75
Qualified "Made in Canada" labels will be used for consumer products that are processed in Canada, but contain imported content, such as "Made in Canada from imported contents."	0.89	0.83
The legislation will also establish stiff offences and penalties for violations.	0.56	0.90
We will ensure that any regulation of natural health products balances the protection of Canadians' health and safety with the freedom to choose alternative products.	0.11	0.84
Establishing Tough Laws on Tampering with Gas Pumps We will amend the Weights and Measures Act and the Electricity and Gas Inspection Act to ensure the accuracy of gas pumps and home-heating meters.	0.89	0.95
Once this increase is fully implemented, it will result in savings of more than $400 per year for eligible seniors receiving the full value of the credit.	0.33	0.61
The restored, lower rate will apply to Canadians who were receiving U.S. Social Security payments prior to January 1, 1996.	0.11	0.65
We have extended the Homelessness Partnering Strategy for the next two years and will maintain funding to deal with homelessness at least at the current levels.	0.78	0.79
We have also renewed funding for the Residential Rehabilitation Assistance Program to assist low-income Canadians in renovating their homes and for the Affordable Housing Initiative, which supports the creation of new rental housing units.	0.11	0.18
Improving Aboriginal education is crucial to giving young members of the Aboriginal community the opportunity to succeed.	0.67	0.43
A re-elected Conservative Government will also commit to pursuing bilateral agreements with provinces to address the wrongs of the residential schools era for Aboriginals attending similar schools not covered by the Indian Residential Schools Settlement Agreement.	0.67	0.89

Table 2.1: Disputed statements among experts and crowd coders

Statement	Pledge mean (experts)	Pledge mean (crowd)
We will achieve its target of a 20 per cent reduction this year in reporting requirements for small businesses and will continue to reduce the burden of unnecessary red tape and paperwork on Canada's small business sector.	0.67	0.96
A re-elected Conservative Government will continue to work with the private sector and small business organizations to identify inefficiencies and make permanent a formal process of measuring, reporting and reducing the burden on businesses.	0.56	0.94

Figure 2.3: Comparing the aggregated pledge evaluation by experts and crowd workers



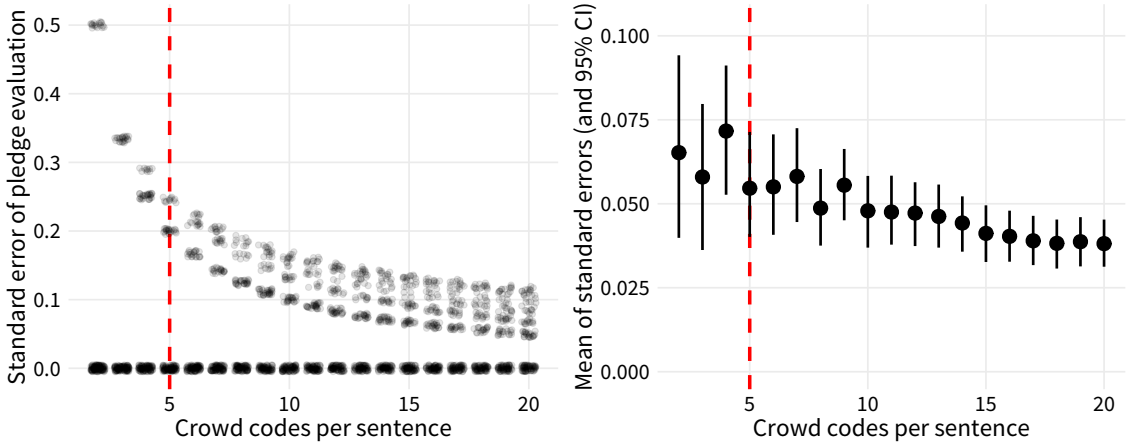
Note: If crowd workers and experts have the same aggregated coding, a point would either be in the bottom-left corner (0%/0%), top-right corner (100%/100%), or close to the diagonal line. The size of the points indicate the number of observations at each location.

I conducted robustness tests relating to the number of required codings, the perceived vagueness of a statement, and the individual characteristics of coders. First, I draw random samples of crowd codings increasing from 2 to 20 for each sentence, and estimate the standard errors in the evaluation of each statement (Benoit et al. 2016). Figure 2.4a shows the standard errors on the level of sentences, showing that a large number of sentences have a standard error of 0: on average, 60 per cent of

the statements are coded in the same way by all crowd workers (which corresponds to a standard error of 0), no matter whether 2 or 20 codings are selected randomly. In other words, even when 20 codings per sentences are selected, for more than half of the sentences all 20 workers made the same coding decision.

Figure 2.4: Standard errors of aggregated pledge evaluations conditional on the number of codes per sentence

(a) Standard error for each sentence **(b)** Mean of standard errors

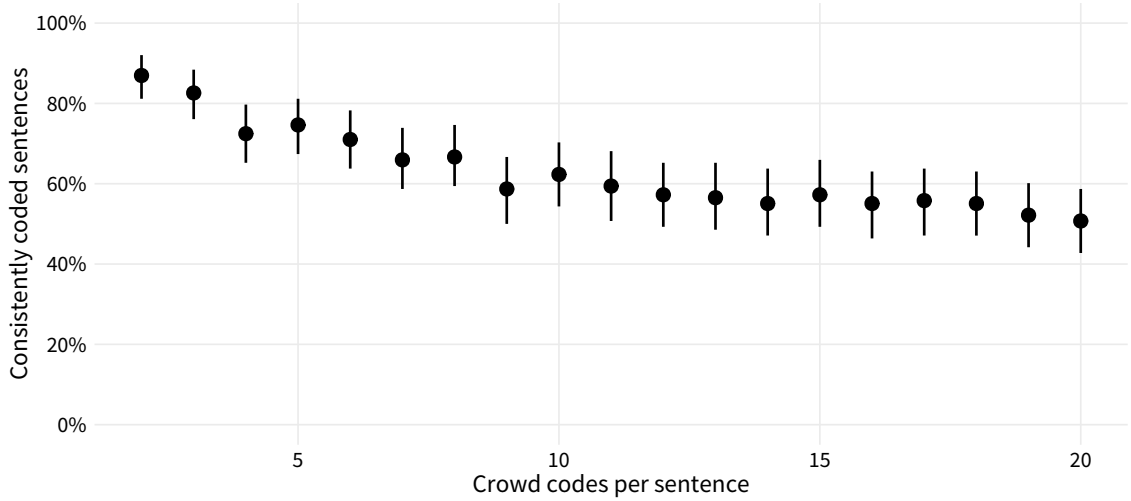


Note: The 'crowd codes per sentence' are retrieved by randomly selecting the respective number of codings from the sample of all judgements. Figure 2.4a plots the standard errors for the aggregated pledge coding for each statement. Each dots marks one sentence. Transparency and random noise are added to reduce overplotting. Figure 2.4b plots the mean of standard errors for the entire sample, along with 95 per cent bootstrapped confidence intervals.

Figure 2.5 visualises the variation in consistently coded statements, conditional on the number of codings. For the 'disputed' statements, the standard errors remain on a rather stable level after collecting at least five codings per sentence. Figure 2.4b plots the means of standard errors conditional on the number of crowd codes, leading to similar conclusions. The mean of all standard errors does not change extensively after five to eight judgements are collected. Thus, as recommended by Benoit et al. (2016), five crowd codings per unit appears to offer reliable aggregated judgements.

Second, I analyse whether the crowd tends to code vague statements as pledges. Therefore, I add varying degrees of the aggregated preciseness score as an additional

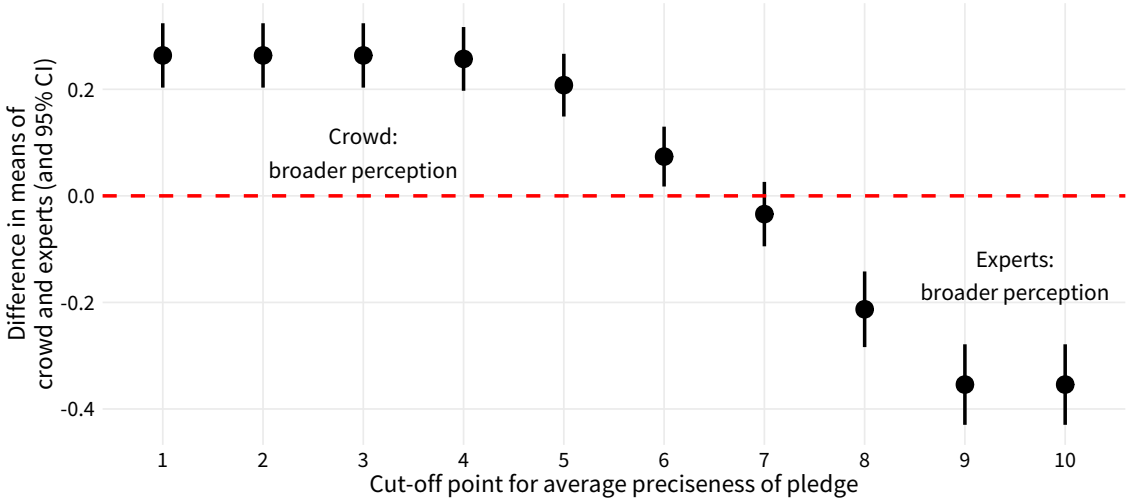
Figure 2.5: Percentage of consistently coded sentences, conditional on the number of coders



Note: The 'crowd codes per sentence' are retrieved by randomly selecting the respective number of codings from the sample of all judgements. 'Consistently coded' implies that all randomly samples judgements were identical (pledge/no pledge). The vertical bars show 95% confidence intervals.

condition for a statement to be considered as a pledge. I expect codings between experts and crowd workers to become more similar with increasing levels of preciseness of a sentence. Figure 2.6 reports the differences in mean pledge evaluations between experts and the crowd based on a paired-sample t-test, and takes the aggregated preciseness into consideration. For instance, a cut-off value of 5 only considers statements as pledges that are coded as a pledge by a majority of the crowd *and* that have an average preciseness above 5. A positive value implies that the crowd has (on average across all aggregated pledge statements) a broader understanding of promises. The significantly and substantively broader understanding by the crowd prevails until a high aggregated value of preciseness (> 7) is used as an additional filter. This finding indicates that the 'testability' criterion of the scholarly definition might not be taken into consideration by many non-experts. The crowd workers' broad understanding could be somewhat balanced out when explicitly taking the proxy of preciseness into account.

Figure 2.6: Comparing the difference in means between aggregated evaluations among experts and crowd workers depending on different cut-offs of preciseness of pledges

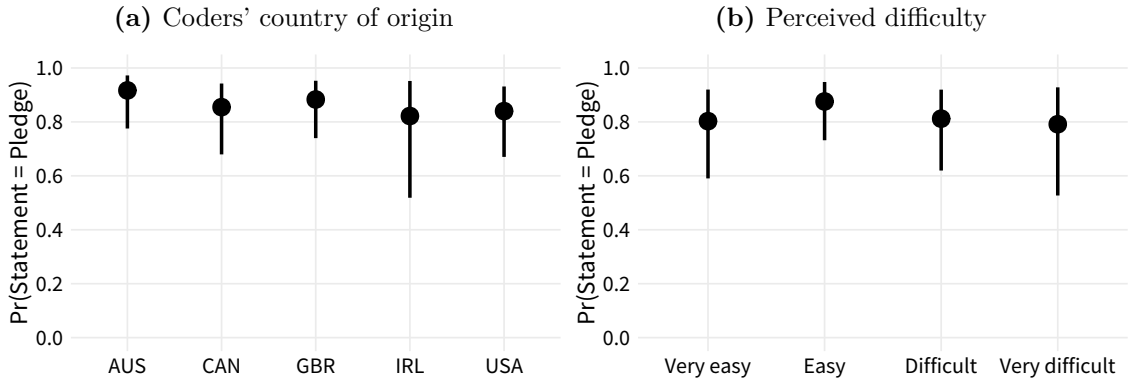


Note: Each dot shows the differences in means based on a two sample t-test comparing experts and crowd workers after taking into consideration the aggregated preciseness rating of a pledge ranging from very vague (1) to very precise (10). For instance, for a cut-off value of 5, only statements that are coded as a pledge by a majority of the crowd *and* have an average preciseness above 5 are indeed considered as a pledge. The bars indicate the 95 per cent confidence intervals of the mean differences.

Third, I test whether the perceived difficulty of coding a statement or the coder's country measured through the IP address (Horn 2018) have an effect on the probability that a statement is coded as a pledge. The binary coding (pledge/no pledge) on the level of individual coders is the dependent variable. The IP address of the coders (on the level of countries), as well as the coder's perceived difficulty of coding each statement, are the independent variables. Random intercepts are included for each statement. The coefficients are small and statistically insignificant. Figures 2.7a and 2.7b plot the predicted probabilities of a statement being coded as a pledge. Neither a coder's country of origin nor the subjective difficulty of coding a statement influence the coding of a statement as a pledge. The test questions throughout the coding job ensured that coders get removed from the exercise (and the subsequent analyses) if they coded future-related, but clearly not testable control

questions incorrectly. The continuous attention checks required coders to stick to the instructions provided before the coding exercise.

Figure 2.7: Predicting the coding of a statement as a pledge



Note: Predicted probabilities are estimated based on the coefficients from Table A.1. The vertical bars show 95% confidence intervals.

Overall, the results allow for three conclusions. First, a considerable proportion of statements are coded as pledge/no pledge by all crowd workers *and* all experts. Second, there is substantial disagreement *within* both groups, and the disagreement among experts is even higher than the disagreement across the crowd. Third, crowd workers classify more statements as pledges, despite following the instructions developed by the group of pledge scholars.

2.4 Discussion

Valid concepts are a crucial condition for research projects across all domains (Adcock and Collier 2001; Gerring 2012). Recently, "[c]rowdsourcing has had a dramatic impact on the speed and scale at which scientific research can be conducted" (Chandler and Shapiro 2016: 53). Building on these advancements, I introduce crowd-coding as a suitable way of reassessing an existing, established concept. The comparison of pledge coding between and within groups of experts and crowd workers reveals that

there is no 'ground truth' in what constitutes an election pledge. Both established scholars and instructed non-experts struggle to code election pledges reliably.

Even though the non-experts were required and continuously monitored in order to code strictly according to the instructions, the crowd evaluates a lot of subjective, not testable promises as election pledges. This finding has serious consequences for empirically testing the 'mandate model of democracy' (McDonald and Budge 2005). The narrower definition of pledges by scholars might explain the perceived gap in pledge fulfilment between experts and citizens (see recently, Thomson 2011; Naurin and Oscarsson 2017; Thomson and Brandenburg 2018; Belchior 2018; Duval and Pétry 2018a). Due to the strict testability criterion, scholars (can) include only a subset of promises in their analyses. Far-reaching, ambitious and often 'exaggerating' (but publicly reported) promises are simply not included in their samples of promises. Yet, voters are likely to perceive such statements as promises. Party manifestos, the prevailing source of coding pledges, are rarely read by voters. In contrast, citizens tend to follow media reports or televised leaders' debates. Therefore, studies on election promises should take additional, more visible sources of campaign communication into consideration. Possibly, non-experts have a broader understanding of pledges because these alternative sources contain more vague information on election promises that do not fulfil the scholars' testability criterion. Voters are likely to note these statements than specific promises from manifestos. Thus, the way media report about election promises requires closer attention.

Moreover, the study underscores that crowd-sourced text analysis not only serves as a means to imitate expert coding or to train supervised classifiers, but also that instructed non-experts indirectly (through the analysis of answer patterns) or directly (through feedback and ratings of the coding process) assess and improve the measurement of a concept. When scholars try to form a new concept, apply an existing concept to a new area, or aim to revisit the background concept (Adcock and Collier 2001), crowd coding reveals whether people without domain-specific

knowledge score textual units in similar ways as experts. To be clear, crowdcoding might not be suitable for jobs that require expert knowledge in a specific research area (Marquardt et al. 2017). Yet, this study highlights that the method can indeed be employed to revisit or refine a familiar concept.

More broadly, the study points to the need for filtering potentially ambiguous or disputed statements. However, collecting multiple codings of all textual units might be infeasible for very large corpora, both for financial reasons and due to time constraints. Future studies should therefore compare and validate machine learning algorithms trained through crowd codings of a random subset of texts. Leveraging a small amount of crowd-sourced human effort at an almost unlimited scale could help researchers to automatically label textual units, but also to detect ambiguous cases. After the automated filtering, these observations could be scored by human coders in order to reduce misclassification and improve reliability.

Chapter 3

Prospective and Retrospective Rhetoric: A New Dimension of Party Competition and Campaign Strategies

Abstract

Voters attribute credit and blame to parties, and parties might react to voters' attributions of responsibility in their campaign communication. However, so far, researchers have not analysed the circumstances under which parties focus on 'retrospective' statements about the present and past, and 'prospective' statements with references about the future. Applying a novel measure of 'retrospective' and 'prospective' campaign rhetoric to 569 national party manifestos from nine democracies and 644 manifestos from German regional elections, I find support for most theoretical expectations. Parties devote, on average, around half of their manifestos to descriptions of the past and present. Incumbents use manifestos to claim credit for past achievements, whereas the opposition points to negative developments. The opposition rarely frames the future significantly more negatively than incumbents. Only parties with low office aspiration in regional elections tend to write 'doomsday' manifestos. The results uncover a new dimension of party competition and have important implications for our understanding of representation and responsibility attribution.

3.1 Introduction

During election campaigns, parties expect voters to respond to strategic messages about who is responsible for policy outcomes. A party can either claim credit for past actions, shift the focus from negative developments by engaging in blame avoidance strategies, or attack other competitors by attributing blame (Weaver 1986). Several studies show that voters indeed attribute credit and blame (e.g., Powell and Whitten 1993; Anderson 1995; Grimmer et al. 2012; Johns 2011). Voters' blame attribution can cause defections and changes in vote choice (Marsh and Tilley 2010). Government parties thus face incentives to emphasise their achievements and distract from negative developments. Opposition parties, on the other hand, might follow the reverse strategy and point to the incumbent's failures. Instead of focusing only on policies for the future, parties can also engage in negative campaigning (Lau and Pomper 2002; Brooks and Geer 2007; Elmelund-Præstekær 2010), and emphasise valence issue (Stokes 1963; Adams and Somer-Topcu 2009; Serra 2010; Curini 2015). Thus, it seems reasonable to assume that election campaigns consist of both credit claiming and blame attribution. Surprisingly, we know very little about the circumstances under which parties choose to engage in blame attribution, and how parties frame the future, compared to the present and past (Dolezal et al. 2018). A better understanding of these factors, however, is required to understand non-policy dimensions of election campaigns and party competition.

In this paper, I introduce the concept of retrospective and prospective rhetoric as a new dimension of party competition. Retrospective statements relate to past or present actions and situations. Prospective statements outline future actions, promises or predicted developments. I measure the extent of prospective and retrospective rhetoric in party manifestos, the central documents of campaign communication (Volkens et al. 2013; Däubler 2012; Harmel 2018). Many researchers have used manifestos to determine issue salience (Budge and Farlie 1977; Budge

2015; Greene 2016) or parties' ideological positions (Laver et al. 2003; Slapin and Proksch 2008; Volkens et al. 2013). However, as Dolezal et al. (2012: 869) claim, "the process of manifesto production, enactment, and public reception are not well understood" (see also the special issue edited by Harmel 2018). Until very recently, political scientists knew very little about the way parties decided upon their campaign strategies, but parties clearly structure campaigns in terms of credit and blame. The title of a former manifesto of the Irish party Fianna Fáil (2002), for example, summarises this deliberate combination of retrospective and prospective elements: "A lot to done. More to do." Because campaign communication can influence a party's reputation, credibility, and competence (Fernandez-Vazquez 2014), research on election manifestos should move beyond the coding of issue positions and saliency.

I argue that parties not only differ in terms of policy issue emphasis, but also in their focus on retrospective and prospective communication. Apart from emphasising and competing on issues based on strong performance in the past (Petrocik 1996), I expect that parties vary in their emphasis and framing of the past and present, and the future. All parties, both incumbents and opposition parties, should focus extensively on retrospective evaluations to gain votes. Yet, retrospective communication serves different purposes. While incumbents claim credit, the opposition attributes blame. Turning to prospective statements, I argue that parties with the ambition to govern after the upcoming election frame the future positively (Harmel and Janda 1994; Mansbridge 2003; Schumacher et al. 2015). Only opposition parties with low aspiration and with low chances of success or aspiration of entering government (Schumacher et al. 2015) should be more tempted to run 'doomsday' campaigns that paint a negative image of the future.

Analysing 569 manifestos from national elections in nine developed democracies and 644 manifestos from regional elections in Germany, I find support for most hypotheses. Large parts of manifestos are devoted to a description of the past and the status quo. Both incumbent and opposition parties, on average, devote around

half of their manifesto to an evaluation of the past and present. Opposition parties indeed engage in blame attribution, while incumbents claim credit for past actions and achievements. 'Doomsday scenarios' are rarely outlined by parties in national first-order elections. Only for regional elections do I find that opposition parties with low office aspiration tend to depict the future more negatively than opposition parties with high office aspiration.

The findings have important implications for party competition, vote-seeking strategies, and campaign communication. First, from a purely descriptive perspective, I show that large parts of manifestos serve purposes of credit claiming and blame attribution. Campaign communication in party manifestos clearly includes non-policy features, as assumed in models of valence (e.g., Adams and Merrill III 2009; Serra 2010; Clark 2014). Parties employ the subjective focus on the past and present to underscore differences with competitors, and focus to a considerable extent on their own achievements, other parties' failures, and descriptions of the current situation. Therefore, researchers should not equate all manifesto sentences with policy pledges. However, for instance, Horn and Jensen (2017) and Alexiadou and Hoepfner (2018) treat the entire manifesto as promise-related statements. Similarly, Kosmidis et al. (2018: 2; emphasis added) claim to "estimate the emotional tone of *policy pledges or initiatives*", but consider all the statements from manifestos and speeches as potential pledge content.

Second, the results strengthen the assumptions of promissory representation. This theory of democracy assumes that parties receive a mandate to carry out promises made during an election and that voters hold parties accountable based on their success or failure of fulfilling these promises (e.g., Mansbridge 2003; Grossback et al. 2005; McDonald and Budge 2005). Ideologically moderate incumbent and opposition parties tend to devote a similar proportion of their manifestos to describing the future.

Third, the paper shows that parties not only emphasise policy issues in manifestos, as suggested by saliency theory (Budge 2015), but also try to persuade voters by focusing on their achievements and the failures of their competitor. Survey evidence shows that voters attribute credit and blame to parties (Iyengar 1991; Anderson 1995; Rudolph 2003; Marsh and Tilley 2010). Yet, previous research has not paid much attention to the question of how parties react to attributions of credit and blame when developing a campaign strategy. Revealing robust differences in sentiment in party manifestos underlines that parties not only engage in negative campaigning in visible arenas, such as televised leaders' debates (Maier and Jansen 2017), but also use the official channel of campaign communication to attribute blame and claim credit. Therefore, future research should pay more attention to non-policy dimensions in party manifestos.

3.2 Theory and Hypotheses

In the following section, I develop hypotheses about two dimensions of campaign strategies: the framing of sentences about the past/present, and different strategies of describing the future.

First, I expect to observe differences in *retrospective campaign communication*. I argue that parties emphasise what they plan for the future, but also evaluate the past and present. Voters tend to consider recent economic or political changes when voting retrospectively (Lewis-Beck and Stegmaier 2000; Achen and Bartels 2016). Additionally, citizens' evaluations of parties' pledge fulfilment is often inaccurate (Thomson 2011; Thomson and Brandenburg 2018; Naurin and Oscarsson 2017). Given this lack of knowledge and short-term focus on recent developments, incumbents need to find ways of minimising electoral losses and convincing voters of their achievements in office. Focusing on promises for the future without assessing the current situation might not be a useful vote-maximising strategy. Therefore, I expect that incumbents

use the public attention they receive prior to an election for claiming credit and avoiding blame by shifting the focus from negative developments (Weaver 1986). For this reason, incumbents should include extensive information on the party's view of the past and present in their manifestos.

Opposition parties, on the other hand, should follow a fundamentally different strategy in retrospective campaign statements. In order to signal criticism towards the incumbent, the opposition should attribute blame towards the government and try to catch voters' attention by focusing on the incumbents' failures. This expectation is partially in contrast with the saliency theory of democracy which assumes that parties selectively emphasise or de-emphasise policy issues where the public perceives the party as competent (Petrocik 1996; Bélanger and Meguid 2008; Budge 2015; Seeberg 2017). The saliency approach neglects the temporal perspective of a statement. A party might focus on a certain issue not because it considers itself as a competent manager in this policy area, but rather because it aims to depict the opponent as incompetent. I argue that the opposition does not profit from only focusing on their own strengths. Instead, they should evaluate the current situation and point to failures of the incumbent. This consideration leads to two observable implications. First, party manifestos contain a large proportion of statements with a retrospective focus. Secondly, when looking at the framing of retrospective sections, incumbents engage in credit claiming strategies, while the opposition attributes blame. As assumed by Traber et al. (Forthcoming), credit claiming should result in the usage of more positive words, whereas blame attribution implies a more negative framing of the past and present.

H 1 (Retrospective Sentiment Hypothesis): *Incumbents frame the past and present more positively than opposition parties.*

The second and third hypotheses relate to *prospective campaign communication.* I argue that parties usually outline their future goals in a constructive optimistic

tone. Incumbent and opposition parties do not profit from describing worst-case scenarios in prospective manifesto sections. Instead, parties aim to convey the image of a competent manager for the future, no matter whether they are in government or in opposition. Two recent studies assess emotive rhetoric in manifestos, but use the entire manifestos without distinguishing between retrospective and prospective parts (Crabtree et al. 2018; Kosmidis et al. 2018). Both studies find that incumbents use more positive language than the opposition. I qualify this claim and state that this difference is driven by retrospective evaluations. In prospective statements about the future, incumbency status should not condition expressed sentiment.

H 2 (Prospective Optimism Hypothesis): *Incumbents and opposition parties use similar degrees of positive sentiment in prospective manifesto sections.*

Finally, I argue that *aspiration to office* conditions how opposition parties talk about the future. Parties follow different goals (Harmel and Janda 1994; Müller and Strøm 1999). While I expect all parties to be vote-seeking (i.e., getting elected), some parties might put more effort into seeking office than others and aim to become part of the government. Other parties might not want to trade-off their core values for government participation (Schumacher et al. 2015). Office aspiration is defined as a party's past participation in government. If a party was frequently part of the cabinet, aspiration is expected to be high, i.e., a party will try to get into government again. Parties might have low levels of office aspiration for various reasons. Single-issue parties might prefer to stick to these issues instead of participating in government (Adams et al. 2006). Alternatively, ideologically extreme parties might be excluded from coalition talks and have no option of joining a coalition. In both scenarios, parties do not have high prospects of getting into office after the election. Office aspiration relates closely to Sartori's (1976) notion of 'government potential'. Some parties – even parties with high vote shares – might be categorically excluded from coalition negotiations or do not even aim to get into government. Higher aspiration

for office thus indicates that a party is both willing and able to become a cabinet party.

Incumbent and opposition parties that have held government offices frequently in the past should act as responsible and serious parties that offer constructive policy (Mansbridge 2003; Dolezal et al. 2018). Parties with high ambition to govern should outline promises for the future and point to the positive implications of the party's potential participation in government On the contrary, parties that have never been in government might be unwilling to frame the future positively, but instead run 'doomsday' campaigns highlighting potential threats of a government consisting of the established parties. In other words, instead of emphasising what could be achieved if the party was in government in the upcoming legislative cycle, parties with low ambition might be tempted to criticise what will go wrong if the established parties with government experience remain in office. I do not expect to observe differences for incumbents conditional on office aspiration. Once in office, government parties should remain positive about the future. Hypothesis 3 follows from these considerations.

H 3 (Office Aspiration Hypothesis): *Opposition parties with low levels of office aspiration frame the future more negatively than opposition parties with high levels of office aspiration.*

Table 3.1 summarizes the expected relationships between the temporal perspective, office aspiration, and the chosen campaign strategy. Plus signs imply positive framing, a minus sign indicates a negative assessment. I expect that incumbent parties claim credit for achievements, resulting in positive sentiment in retrospective manifesto sections. Since office aspiration is a continuous measure (see below), there are instances where parties with low or without any level of previous government experience enter a cabinet. I expect these incumbents with (previously) low office aspiration also to be very positive in retrospective and prospective statements. Oppo-

sition parties with low office aspiration should be especially negative in retrospective manifesto sections and remain negative in prospective parts. Opposition parties with high office aspiration should also be negative in retrospective sections, but the sentiment in prospective parts should approximate the rhetoric by incumbents.

Table 3.1: Theoretical expectations about sentiment in prospective and retrospective manifesto sections

Office aspiration	Government status: **Opposition**	Government status: **Incumbent**
Low	Retrospective: − − Prospective: −	Retrospective: ++ Prospective: ++
High	Retrospective: − Prospective: ++	Retrospective: ++ Prospective: ++

3.3 Data and Case Selection

Party manifestos contain the definitive statement of a party's officially reported position (Budge et al. 2001). Parties consider very carefully how to structure and what to include in an election manifesto (Däubler 2012). A growing body of work shows that manifestos are crucial documents to promote intra-party unity, and that these texts are taken seriously by parties and noticed by voters. Candidates use manifestos to inform voters about their party's positions on a variety of issues (Eder et al. 2017), and media report frequently about manifestos before elections (Bischof and Senninger 2018). Using manifestos as a proxy for campaign strategies should constitute a least likely case to observe differences in sentiment and retrospective assessments. Manifestos are drafted carefully in a lengthy process (Dolezal et al. 2012), and saliency theory assumes that parties do not mention other parties, but rather emphasize certain topics more than others (Budge 2015). More 'confrontational'

campaign tools such as campaign speeches or leaders' TV debates should reveal stronger effects in terms of retrospective focus and sentiment (Brooks and Geer 2007; Maier and Jansen 2017; Hopmann et al. 2018).[1] If we find support for the hypotheses in manifestos, the effects should persist or increase in other forms of campaign messages. I expect the hypotheses to hold regardless of the cultural or institutional environment. Therefore, I analyse manifestos from first-order national elections across nine countries, and from subnational election manifestos across the 16 German Länder. The focus on the regional level has the advantage that institutional and cultural environment can be held almost constant. Without facing the problem of spatial and cultural differences, the subnational comparative method allows to focus on the central explanatory variables and minimises confounding effects caused by cross-country differences (Snyder 2001; Bowler et al. 2016). At the same time, external validity of the findings should increase if we observe similar results across countries.

3.3.1 National General Elections

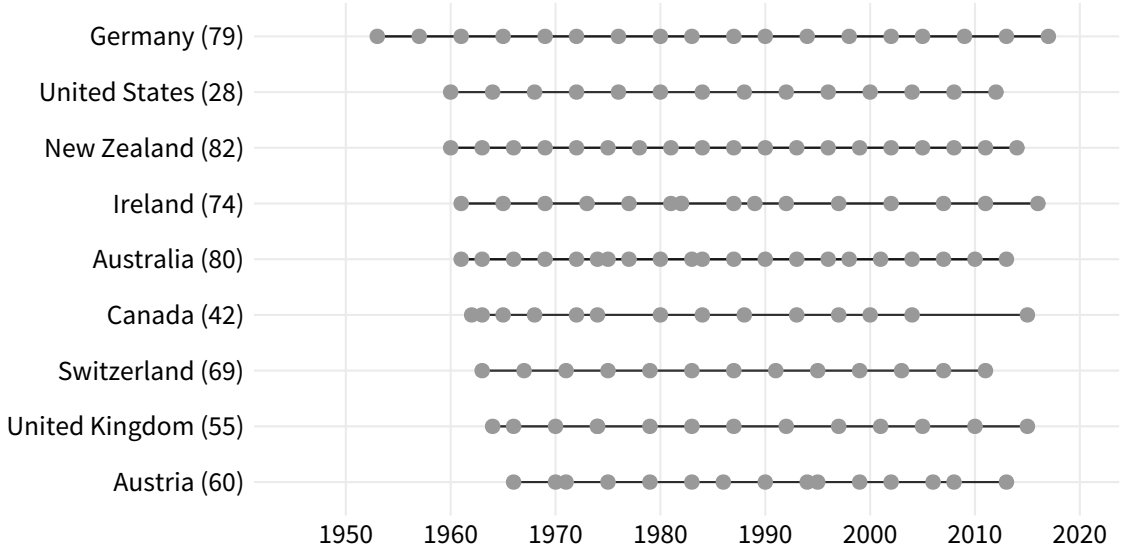
The analysis includes all available machine-readable English and German manifestos from developed democracies (Austria, Australia, Canada, Germany, Ireland, New Zealand, Switzerland, the United Kingdom, and the United States) resulting in a sample of 569 manifestos published between 1953 and 2017 consisting of 392,239 sentences.[2] The number of elections per country ranges from 13 (United Kingdom) to 21 (Australia), and the number of manifestos per country from 28 for the United

[1] Kosmidis et al. (2018) show that US presidential State of the Union Addresses contain more affective emotions than manifestos.

[2] These machine-readable party manifestos are available in the recently published Manifesto Corpus with the `manifestoR` R package (Merz et al. 2016; Lehmann et al. 2017). Note that I use quasi-sentences as the unit of observation for annotated manifestos (Däubler et al. 2012). Reshaping from quasi-sentences to natural sentences is problematic as some manifestos do not contain the original punctuation which results in the erroneous combination of several natural sentences to one observation. Including a dummy variable for annotated manifestos to control for unobserved effects due to the slightly different units of analysis does not change any of the reported results.

States to 82 for New Zealand. Figure 3.1 summarises the time frames for each country.

Figure 3.1: Available machine-readable national party manifestos



Note: Each dot marks one election. Number of manifestos per country in parentheses. The first German federal election in 1949 is excluded from the analysis because of missing information on incumbency status.

The cross-country comparison maximises variation in institutional variables (Powell 2000; Lijphart 2012). The sample of countries consists of presidential, parliamentary, and mixed-systems, applying a variety of institutional rules and variation in the effective number of parties. The United Kingdom is a unitary country that employs first-past-the-post in single-member districts. The United States is a presidential system. Switzerland is neither parliamentary nor presidential (Strøm 2000: 266) and most parties participate in government. Austria employs proportional representation (PR), and Germany uses mixed-member PR. Until 1995, elections in New Zealand were conducted under first-past-the-post. Afterwards, the country switched to mixed-member PR. The Republic of Ireland is a unitary country using a single-transferable vote system, resulting in a strong focus on constituency service. Elections in Australia are conducted under the alternative vote system. If we find support for the hypotheses across these diverse settings, we should have confidence

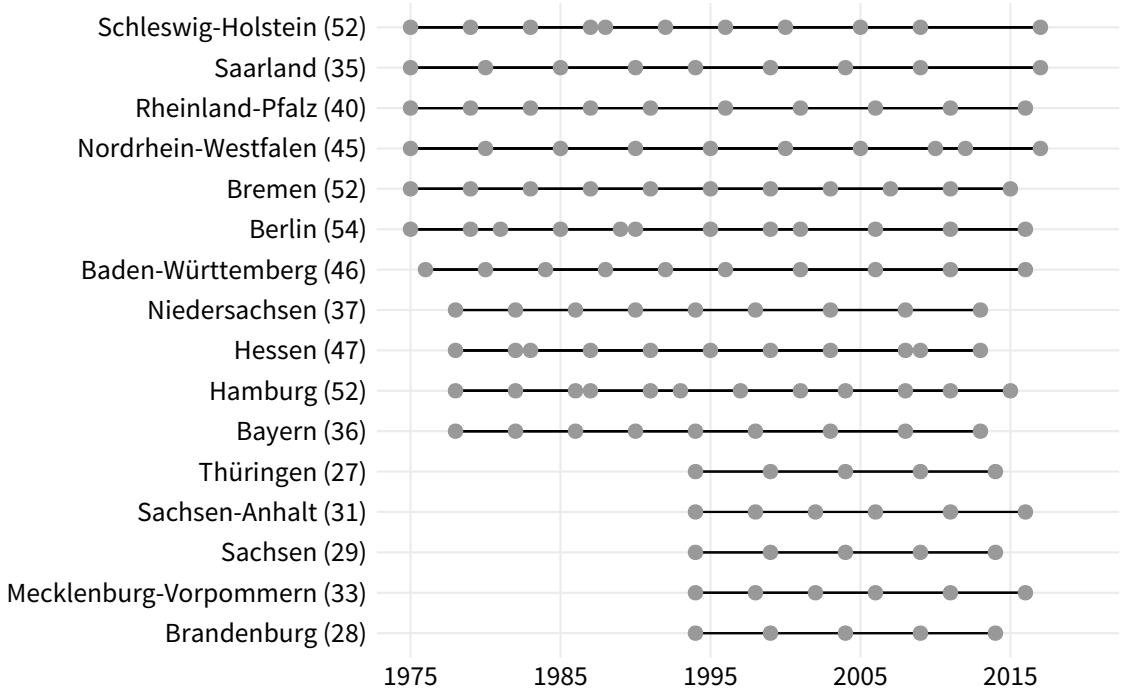
that the findings are not driven by specific institutional rules. Clearly, it would be desirable to consider additional language to assess, for instance, how manifestos written in futureless tongues employ retrospective and prospective statements (Pérez and Tavits 2017). Nonetheless, manifestos from 144 elections over a period of over 60 years provide variation in all key independent variables of interest for this study.

3.3.2 Subnational Elections

The focus on the subnational level allows to focus on the central independent variables of interest: incumbency status and office aspiration. I reproduce the cross-country analysis with 644 party manifestos from 141 German Länder elections in the period between 1970 and 2017. Although the power of the subnational governments is limited to a small number of policy issues and even though Land elections are considered as second-order elections (Jeffery and Hough 2001), the government of the 16 Länder constitute the Bundesrat, Germany's upper chamber. Based on the documents available on *Political Documents Archive* (Benoit et al. 2009a; Gross and Debus 2018), I constructed a text corpus consisting of 576,187 sentences from these manifestos. Figure 3.2 plots the availability of manifestos for the 16 Länder. For each of the 'old' West German states, manifestos of the major parties are available since the 1970s. For the 'new' East German Länder, I include manifestos since 1994.[3] It is important to note that German subnational party factions write custom manifestos for each election and usually do not copy previous manifestos or federal party manifestos. By estimating document similarities between all manifestos, in Section B.4, I show that only 10 out of the 644 documents have been reused. Excluding these manifestos from the analysis does not change any of the conclusion.

[3]The documents from the elections in 1990 in the 'new' Länder are excluded because of missing information on the incumbency status. Results do not change when including the 1990 elections, thus increasing the corpus to 679 manifestos and coding all parties from inaugural elections in the East German states as opposition parties.

Figure 3.2: Available machine-readable subnational party manifestos



Note: Each dot marks one election. Number of manifestos per Land in parentheses. The first election in the 'new' states that joined Germany after re-unification in 1990 are excluded from the analysis because of missing information on incumbency status.

3.4 Measurement

In this section I outline how I measure the temporal focus of a statement, how I conceptualise blame attribution and credit claiming, and summarise the measurement and sources of the independent and control variables.

3.4.1 Classifying Retrospective and Prospective Statements

I define statements about the past and present as 'retrospective', whereas statements with a reference to the future are categorized as 'prospective'. Many sentences refer to the past and the present situation. Classifying sentences relating to the past and present as 'retrospective' seems reasonable based on the assumptions of retrospective voting (Healy and Malhotra 2013). Due to cognitive biases, voters remember what has happened in the recent past or take the current situation into account when

casting a retrospective vote (Achen and Bartels 2016). Talking about the present situation therefore constitutes retrospective campaign rhetoric. In other words, as long as a sentence does not include any information on the future, I classify the statement as retrospective (Dolezal et al. 2018).

The novel approach of distinguishing between retrospective and prospective statements requires a valid and reliable classification. Detecting the temporal dimension on the level of sentences is difficult for several reasons. First, the traditional manual coding is expensive in terms of time and resources (Hopkins and King 2010; Benoit et al. 2016). Secondly, fully automated methods based on off-the-shelf dictionaries perform poorly for tense classification.[4] Selecting keywords manually and creating a dictionary based on these terms often captures only a subset of the relevant words which also results in poor classification of unknown texts (King et al. 2017). Part-of-Speech (POS) taggers are very good at annotating the part of speech to individual terms, but fail in detecting the tense of an entire sentence.[5]

In the absence of a reliable existing method, I classify prospective and retrospective sentences with Naïve Bayes, a supervised machine learning algorithm (Manning et al. 2008). Generally speaking, human coders train a classifier by labelling a subset of the text corpus with the correct coding. The class of each document (i.e., sentence) in this training corpus is thus known a priori. Afterwards, the classifier is applied to an unlabelled set of sentences. Based on a comparison of word frequencies in the training and test corpus, the classifier assigns posterior probabilities for a sentence belonging to each class.

In order to create a training set for the Naïve Bayes classifier, I chose both crowd-sourced text coding (Benoit et al. 2016) and the existing tense classification of party manifestos provided by the Austrian National Election Study (Müller et al.

[4]The existing 'tense dictionaries' provided in the 2001 and 2007 versions of the frequently used LIWC dictionary (Tausczik and Pennebaker 2010) were removed in the most recent release due to unreliable results (Pennebaker et al. 2015: 7).

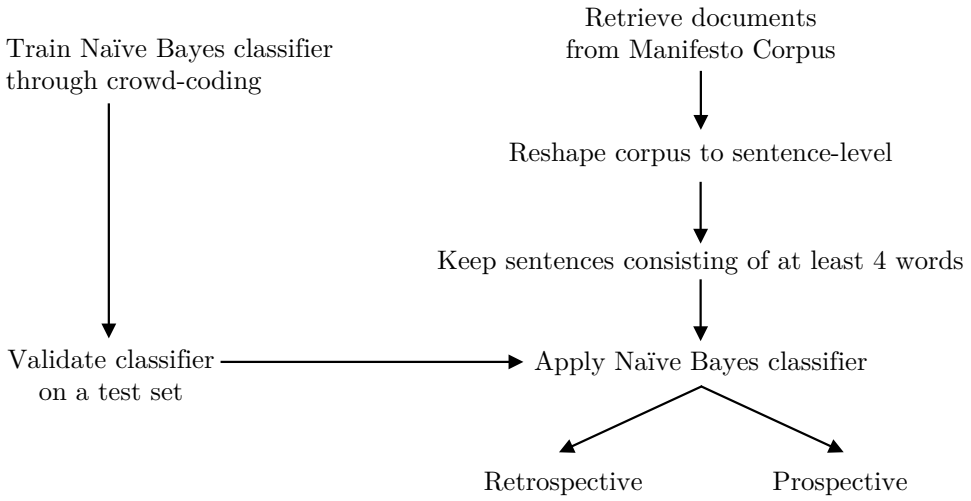
[5]Careful attempts to use the `spaCy` POS tagger for this type of annotation did not work reliably (F1 scores between 0.55 and 0.70).

2017). The coding of the Austrian manifestos is carried out by trained research assistants and the tense of each sentence is scored by two coders (Dolezal et al. 2016). Crowd coding has been employed in several studies to label sentences from party platforms (e.g., Benoit et al. 2016; Horn 2018). These aggregated non-expert judgements proved to be a reliable alternative to expert coding when the workers were provided with clear instructions and when comprehensive quality checks are added to the coding job. The English sentences used for training the classifier were coded by at least five instructed non-experts. I employed strict quality standards using a entry quiz, and randomly included 'gold questions' throughout the coding job. One out of five sentences is a 'gold question' with a predefined answer. These gold questions are sentences that are clearly future or present/past orientated. The coding of the gold questions comes from an expert reliability test conducted for the comparative study on pledge fulfilment by Thomson et al. (2017). Throughout the job, coders have to answer at least 80 per cent of the 'gold questions' correctly. For more details on the crowd-sourced approach, text selection and the quality control mechanisms see Sections B.1 and B.2.

Having created the training corpora, all sentences from the training and test set are converted into a document-feature matrix where each sentence constitutes one document. Terms are changed to lowercase, and punctuation is removed, but I do *not* remove any 'stopwords' or stem terms as the ending of verbs are crucial to distinguish between tenses. If the posterior probability of 'prospective' exceeds the threshold of 0.5, the sentence is coded as 'prospective'. Posterior probabilities below 0.5 indicate a 'retrospective' statement. This threshold is typical for the classification of binary categories (Manning et al. 2008). Figure 3.3 visualises the process of retrieving and reshaping the manifesto texts as well as training and applying the classifier.

Figure 3.4 offers examples of the classification of typical German and English manifesto sentences in various tenses and with context-dependent meanings of words. A longer horizontal bar indicates a higher estimated posterior probability of a

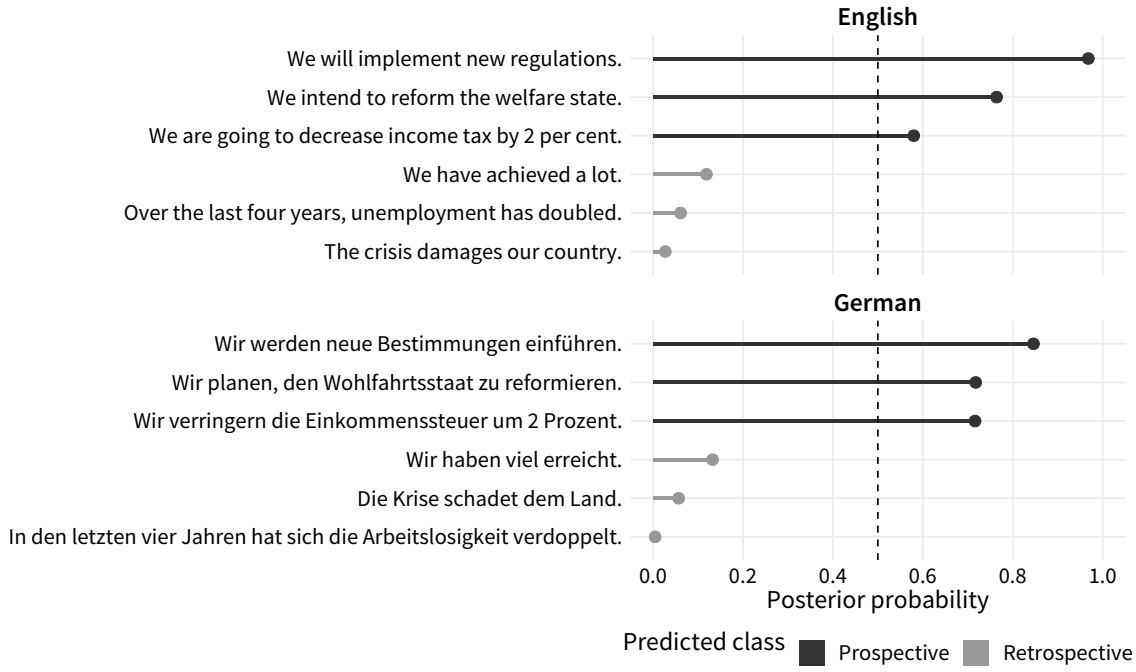
Figure 3.3: Overview of the methodological approach for the classification of manifesto sentences



sentence belonging to the 'prospective' class. Even the sentences that are written in the present tense, but relate to the future are classified correctly. Figure B.10 shows 'keyness' statistics which are scores for features that occur differentially across sentences classified as retrospective and prospective (Bondi and Scott 2010). Words such as 'will', 'new', 'ensure', 'provide' and 'continue' are far more likely to occur in the prospective category, while terms like 'has', 'been', 'now', 'since' have a much higher probability of not being present in prospective statements.

Validation is essential when working with supervised or unsupervised quantitative text analysis (Grimmer and Stewart 2013). I compare the English and German classifications with a randomly sampled held-out test set of 300 sentences in each language (see extensively Section B.6). For the English texts, precision, recall and the F1 score (a harmonic mean of recall and precision) range between 0.85 and 0.9 for Naïve Bayes. The German classification does not work as reliably (F1=0.76), but Naïve Bayes produces the best classification results compared to other classification

Figure 3.4: Assessing the face validity of the Naïve Bayes classification for typical English and German manifesto sentences



Note: The x-axis displays the estimated probability for each sentence belonging to the class 'Prospective'. A posterior probability above 0.5 for 'Prospective' (vertical dotted line) is used as the benchmark for the final categorization.

methods.[6] The small degree of misclassification will allocate some positive prospective sentences to the retrospective category, and negative retrospective statements to the prospective section. As a result, the obtained classification should provide conservative estimates.

3.4.2 Measuring Sentiment

The next challenge relates to the measurement of the dependent variable. In retrospective parts, claiming credit for achievements should imply the use of more positive words, while blame attribution should increase the number of negative terms (Traber

[6]I also tested as to whether the F1 score decreases if I predict the classes of manifestos published in earlier times as the training set. In an additional exercise I crowd coded 50–150 randomly sampled sentences from the party platform of the Conservative Party of Canada (2008) and the manifesto of New Zealand's Labour Party from 1960. The F1 scores are remarkably stable over time and across countries (F1 ranging between 0.89 and 0.92 for the three manifesto samples).

et al. Forthcoming). Therefore, a measure of sentiment taking into consideration positive and negative terms should capture the expectations relating to campaign strategies in prospective and retrospective manifesto sections.

Following recent studies on emotive rhetoric (Crabtree et al. 2018; Kosmidis et al. 2018; Proksch et al. 2019) and blame avoidance in political texts (Traber et al. Forthcoming), I measure the extent of blame attribution and credit claiming with a dictionary approach. The Lexicoder Sentiment Dictionary (LSD) (Young and Soroka 2012) and the Linguistic Inquiry and Word Count (LIWC) (Tausczik and Pennebaker 2010) are two of the most frequently used sentiment dictionaries. The LSD has been developed for the analysis of economic newspaper articles (Soroka 2012). Recently, this dictionary has also been successfully applied and validated for political speech (Proksch et al. 2019). Proksch et al. (2019) translated the English LSD to other languages, and show that the English and German dictionaries perform particularly well compared with hand coding of perceived sentiment ($r > 0.8$ for German and $r > 0.85$ for English).

Since the LSD outperforms LIWC – developed by psycholinguists – in terms of face validity (Soroka et al. 2015), I use the Lexicoder as the main dictionary. Note, however, that the correlations for sentiment on the level of manifestos are high and that the results remain unchanged when I use the LIWC dictionary instead. Table 3.2 lists the number of positive and negative terms in each dictionary.[7] The ratio of positive to negative indicates that each collection of words contains more negative than positive words.

Following Proksch et al. (2019), I measure *Positive sentiment* as the logged ratio of positive to negative terms in a sentence (see also Lowe et al. 2011):

$$Sentiment = log\left(\frac{\sum positive + 0.5}{\sum negative + 0.5}\right)$$

[7]The English version of the LSD contains negations, whereas negations are missing in the LIWC dictionary. I count negated positive (negative) words as negative (positive).

Table 3.2: Overview of English and German sentiment dictionaries

Dictionary	Positive terms	Negative terms	Positive/Negative
LSD (English)	1,709	2,858	0.60
LIWC (English)	407	500	0.81
LSD (German)	2,287	4,595	0.50
LIWC (German)	643	1,043	0.62

A positive value indicates that a sentence contains more positive than negative terms. If the number of positive terms equals the number of negative words (or if a sentence does not contain any words from the sentiment dictionary), *Positive sentiment* takes the value 0. As an additional robustness check, I also conceptualise sentiment as the difference between positive and negative words divided by all words (Soroka 2006; Crabtree et al. 2018; Rauh 2018).[8] The alternative aggregation leads to the same or even stronger results. Having classified each sentence as either prospective or retrospective, I average the net positive sentiment (a) for the entire manifesto, for (b) retrospective parts, and (c) prospective sentences. Tables B.1–B.4 provide examples of retrospective and prospective sentences from English and German manifestos with very positive and negative sentiment scores.

3.4.3 Independent and Control Variables

The three key independent variables are *Incumbency status*, the *Temporal dimension*, and *Office ambition*. *Incumbency status* is measured as a binary variable, indicating whether a party was an incumbent or opposition party at the time when a manifesto was drafted. Since I do not have theoretical reasons to expect variation within different types of incumbents or the opposition, I opt for the binary classification. A more fine-grained differentiation that distinguishes between Prime Ministerial party, coalition partners, opposition parties represented in parliament, and opposition parties not represented in parliament do not change the substantive conclusions. For the national elections, I use the ParlGov dataset (Döring and Manow 2018) to code

[8] This measure estimates sentiment as: $100 \times \frac{\sum positive - \sum negative}{\sum words}$.

incumbency status based on the last cabinet before the upcoming election. Data were added manually for recent elections and the United States. I use the data from (Reck and Weber 2010) to code the incumbency status for parties in the German Länder until 2010 and added the cabinets for all elections between 2010 and 2017. I argue that sentiment in retrospective parts depends on the incumbency status. I therefore add an interaction effect between *Incumbency status* and the *Temporal dimension*.

I follow Schumacher et al.'s (2015) measurement of *Office aspiration*. The continuous measure can range between 0 (a party has never been in government prior to election t) to 1 (a party has always been in government).[9] The measurement of office aspiration has the advantages of being updated after each election. I select the last cabinet prior before the upcoming general election as the indicator for government participation.[10] The United States are excluded from the analysis regarding office aspiration because the measure only applies for parliamentary democracies. Both for the federal and subnational elections, the distribution of Office aspiration is skewed towards 0 because many parties never held government office.

Turning to the control variables, I first add measures of left-right ideology to the regression models. Ideologically extreme parties are more likely to be in opposition and therefore more likely to employ negative sentiment in prospective manifesto sections. I include *Left-right* and *Left-right*^{2} since I expect a curvilinear relationship with more extreme parties using more negative sentiment. I include two measures for left-right ideology (Laver 2014). For national elections, I opt for the left-right scale (RILE) developed by the Comparative Manifesto Project (Budge and Meyer

[9]The formula for *Office aspiration* considers office performance (X) which takes the value 1 (0) if a party was in government (opposition) in the last cabinet of a cycle. *Office aspiration* is then calculated as: $(X_{t-1} + X_{t-2} + ... + X_{t-n})/n$, where n is the number of cabinets a party could theoretically have been part of.

[10]The last cabinet of a cycle appears as the most suitable measure for this paper since I expect parties to draft manifestos based on their government status right before an election. Results do not change if I use the first cabinet of a cycle instead: cabinet reshuffles within a legislative cycle did not occur frequently in the countries under investigation.

2013). For the subnational elections, I locate texts on a unidimensional scale with the unsupervised Wordfish scaling method (Slapin and Proksch 2008; Lowe and Benoit 2013). Wordfish estimates serve as the preferred measure for the subnational level because regional parties are not part of the Comparative Manifesto Project and because party positions on the federal level do not always correspond to the regional party's position (Bäck et al. 2016).

Before scaling the texts, I subset corpus by decade and separating the corpus into 'old' and 'new' Länder.[11] The boxplots in Figure B.4 show the distributions of ideological positions for all parties. The Green party and the Left have the lowest median of theta (θ) values, indicating leftist positions on a unidimensional ideological scale. As expected, the SPD, CSU, and CDU are the most centrist parties. The FDP is positioned to the right of the CDU, whereas manifestos from the AfD and NPD have mostly large and positive values.[12] The face validity of these positions justifies the usage of Wordfish to scale subnational party manifestos.

Attention to prospective statements as well as variation in sentiment and issue diversity in retrospective parts could be influenced by the economic situation. If the economy performs well in the year of an election, voters usually attribute this situation to the incumbent (e.g., Lewis-Beck and Stegmaier 2000; Duch and Stevenson 2008; Achen and Bartels 2016). Previous findings suggest that the economy impacts manifesto sentiment and concreteness (Crabtree et al. 2018; Eichorst and Lin 2019), pledge fulfilment (Thomson and Costello 2016) and issue diversity (Greene 2016). Economic data on *GDP growth* and *Inflation* were derived from the World Bank. For the German regional level, I estimate the *Change in the unemployment rate*

[11] I run Wordfish separately on these subsets for two reasons. First, party competition in the 'old' and 'new' German Länder might differ fundamentally. For instance, the Left party (formerly PDS) differs both in terms of positions and electoral support in old and new states. Second, a separation of manifestos by Bundesland or without filtering based on decades results in lower face validity of the estimates.

[12] Only the *positive* θ values of the Pirate party contradict their actual ideological position, which is assumed to be on the left side of the political spectrum. Changing the positive values for the party to negative ones by multiplying θ by –1 does not make any differences in the results, since I assume curvilinear effects for party ideology.

compared to the previous year in each Land. As governments are expected to react to earlier economic developments (Crabtree et al. 2018), I lag the economic variables by one year. To control for the possibility that negativity in election campaigns might have increased in recent years (Geer 2012), I add the continuous variable *Year* to control for changes over time.

Since the dependent variable *Sentiment* is an approximately normally distributed continuous measure, I run multilevel linear regression models (Bates et al. 2015). I opt for multilevel models because I expect the observations to be clustered into countries/Bundesländer, parties, and elections. Therefore, I add random intercepts for each country/Bundesland, party, and election. I also add random intercepts for manifestos, since each document occurs twice in the dataset (once for the value of prospective and once for the retrospective sentiment). Results remain unchanged if I run separate regressions for the retrospective and prospective subsets instead of interacting *Incumbency status* and the *Temporal dimension*. Since *Office aspiration* and the party show high multicollinearity – some parties are always in opposition – I remove random intercepts on the party level when I add the measure of aspiration to the regression models.

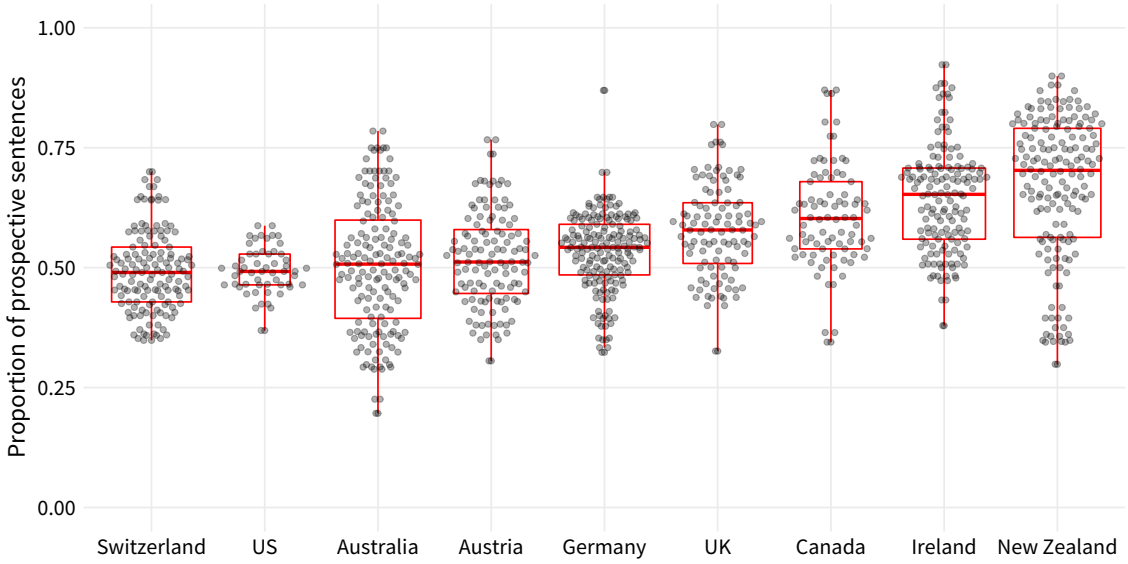
3.5 Results

In this section, I report and discuss the results from the empirical analysis. First, I provide descriptive evidence about the variation and extent of retrospective proportions in party manifestos. Second, I test Hypothesis 1 relating to credit claiming and blame attribution in retrospective manifesto sections. Afterwards, I report the results relating to Hypothesis 2 which posits that incumbents and oppositions do not differ significantly in terms of sentiment in prospective manifesto sections. Finally, I provide evidence as to whether (opposition) parties with lower office aspiration tend to draft 'doomsday' manifestos (Hypothesis 3).

3.5.1 Retrospective and Prospective Campaign Communication

From a purely descriptive perspective, I first offer evidence about the proportion of retrospective statements in party manifestos. Figure 3.5 displays the distribution of prospective sentences divided by the total number of sentences for 569 English and German manifestos from national elections. Across all countries, between one third to two thirds of manifesto sentences relate solely to the past or present. The median of prospective statements in Switzerland, Austria, Australia, Germany and the United States lies at around 0.5 meaning that one in two sentences relates to the past or present. Ranging at around 70 per cent, manifestos published in New Zealand, Ireland and Canada have the highest median of prospective statements. With a mean of 0.5 (SD = 0.07) and a range between 0.3 and 0.76, prospective rhetoric in German subnational party manifestos follows a remarkably similar distribution.

Figure 3.5: Proportions of prospective sentences across manifestos

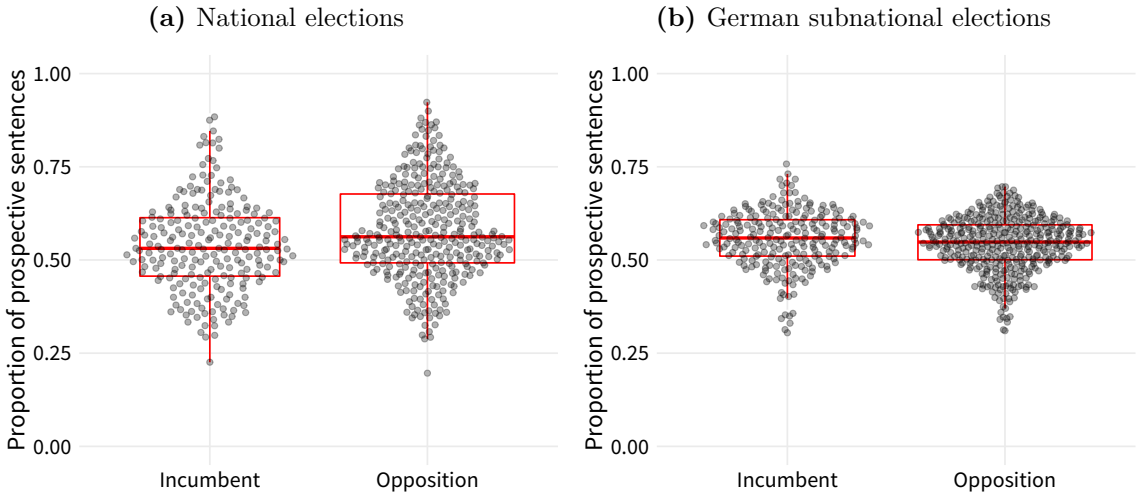


Note: Each dot represents one of the 569 party manifestos included in the analysis.

Figure 3.6 plots the proportion of prospective statements for incumbent and opposition parties. While the median of prospective emphasis is slightly larger for opposition parties than for incumbents in general elections (around 3 percentage

points), we do not observe any substantial differences based on incumbency status in the German regional elections. Incumbents and opposition parties put considerable emphasis on retrospective statements in first- and second-order elections.

Figure 3.6: Proportions of prospective sentences faceted by incumbency status



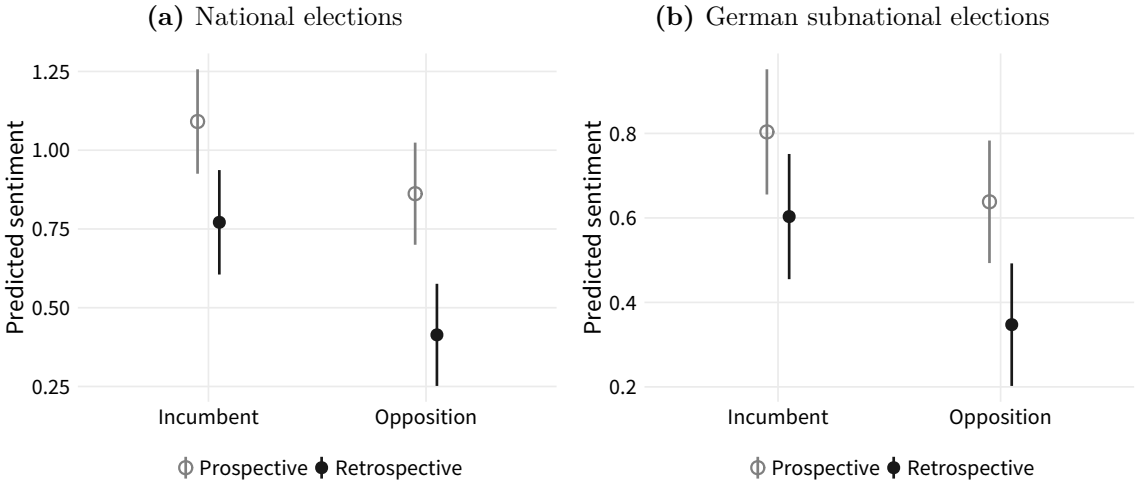
3.5.2 Sentiment in Prospective and Retrospective Statements

Having revealed considerable variation in the amount of prospective salience across manifestos, I analyse how sentiment differs between prospective and retrospective manifesto sections. Figure 3.7 plots the predicted values of sentiment for the national elections (left-hand panel) and German subnational elections (right-hand panel). First, we observe that the net sentiment is positive both for incumbent and opposition parties. Even parties in opposition use more positive than negative terms. This finding mirrors the results from previous studies on emotive rhetoric by parties (Crabtree et al. 2018; Kosmidis et al. 2018; Proksch et al. 2019) and corresponds to the universal positivity bias observed across many languages (Dodds et al. 2015).

Differences in sentiment between retrospective and prospective statements are considerably smaller for incumbents than for opposition parties. The statistically

significant and sizeable interaction effect of *Opposition* and *Retrospective* across all Models of Table 3.3 shows that opposition parties are more negative in their assessment of the past and present. Opposition parties' difference in prospective and retrospective sentiment amount to 0.45 (national) and 0.3 (subnational elections). These values correspond to 93 and 75 per cent of the standard deviations of sentiment in each sample, indicating a large difference for opposition parties. For incumbents, the difference amounts to 66 and 51 per cent of the dependent variables' standard deviations. These findings offer strong support for Hypotheses 1 and 2. Incumbents frame only the past and present in a significantly more positive way than the future.

Figure 3.7: The moderating impact of the temporal focus on manifesto sentiment



Note: The y-axis shows the predicted net positive sentiment in a manifesto, the vertical lines display 95 per cent confidence intervals (based on Models 1 and 3 in Table 3.3).

Turning to the control variables, I find significant effects of party size (measured as the *Seat share* at the election) only for the German subnational elections. This finding is likely caused by the higher number of small and regional parties that are not represented in parliament but employ very negative sentiment. Changes in the economy in the year prior to an election do not have a significant influence on manifesto sentiment, and the inclusion of the economic variables does not change the size of the remaining coefficients (Models 2 and 4 of Table 3.3). Figure B.6 shows

that, on average, ideologically right parties in national elections tend to express more negative sentiment than moderate parties. In the German subnational elections, we observe a curvilinear relationship with ideologically more extreme parties on the left and right being less positive. Yet, in both cases the substantial differences based on ideology are considerably smaller than the impact of incumbency status.

Table 3.3: Predicting sentiment in party manifestos

	M1 (Nat.)	M2 (Nat.)	M3 (Subn.)	M4 (Subn.)
Opposition	-0.23^{***}	-0.23^{***}	-0.17^{***}	-0.17^{***}
	(0.03)	(0.03)	(0.03)	(0.03)
Class: Retrospective	-0.32^{***}	-0.33^{***}	-0.20^{***}	-0.20^{***}
	(0.03)	(0.03)	(0.02)	(0.02)
Seat share	0.11	0.06	0.34^{**}	0.32^{**}
	(0.12)	(0.13)	(0.12)	(0.12)
Left-right	-1.83^{**}	-2.12^{***}	0.85^{*}	0.78
	(0.56)	(0.57)	(0.42)	(0.44)
Left-right^2	-0.95^{*}	-0.90^{*}	-1.66^{***}	-1.66^{***}
	(0.44)	(0.45)	(0.36)	(0.37)
Year	0.00	0.00^{*}	0.01^{***}	0.01^{***}
	(0.00)	(0.00)	(0.00)	(0.00)
Opposition × Class: Retrospective	-0.13^{***}	-0.12^{***}	-0.09^{***}	-0.09^{***}
	(0.03)	(0.03)	(0.02)	(0.02)
Change in GDP growth (lag)		−0.00		
		(0.00)		
Change in unemployment rate (lag)				−0.01
				(0.01)
AIC	822.51	699.01	74.72	103.16
BIC	887.99	767.72	141.81	174.95
Log Likelihood	-398.25	-335.50	-24.36	-37.58
N	1138	1000	1288	1246
Num. groups: Manifesto	569	500	644	623
Num. groups: Election	144	122	141	138
Num. groups: Party	97	94	15	15
Num. groups: Country	9	9		
Num. groups: Bundesland			16	16

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

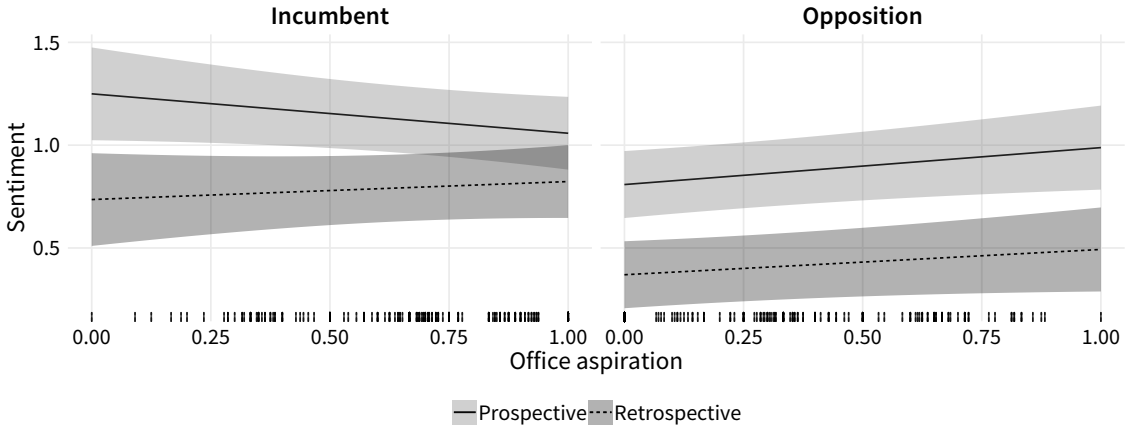
Note: Multilevel linear regressions with the sentiment as the dependent variable. Models 1 and 2 focus on national elections, Models 3 and 4 on German subnational elections. Models 2 and 4 reproduce Models 1 and 3, but add economic indicators. Standard errors in parentheses.

3.5.3 The Impact of Office Aspiration on Campaign Sentiment

Hypothesis 3 posits that office aspiration conditions opposition' parties sentiment. To test this conditional hypothesis, I add a three-way interaction between *Predicted class*, *Incumbency status* and *Office aspiration* to the regression models. For the national elections, we do not observe the expected interaction. Although the direction of the relationship is positive, the coefficient is small and the confidence intervals are very wide. Opposition parties with low office aspiration in national elections are not significantly more negative in their assessment of the future (Figure 3.8 based on Model 1 of Table 3.4). However, we observe a sizeable effect of aspiration for office in the case of regional elections (Figure 3.9). Looking at incumbent parties first (left-hand panel), we see only a very small increase in sentiment when incumbents have a long history of being in office. For opposition parties, on the other hand, we observe a much stronger effect. Parties that have never been in government are on average over 2.5 times more negative in their assessment of the future than opposition parties that have been represented in 80 per cent of the previous cabinets. This increase in sentiment occurs also in retrospective sections. Opposition parties with high office aspiration engage less in blame attribution than opposition parties with low office aspiration. Regional factions of German parties with low office aspiration are more tempted to draft 'doomsday' manifestos that describe the future very negatively.

A possible explanation for the strong impact of office aspiration on the regional, but not on the national level, relates to the nature of subnational elections. Regional elections are frequently regarded as second-order elections used by voters to 'punish' the federal government (Schmitt and Teperoglou 2017). The federal government often loses votes in regional elections (Jeffery and Hough 2001; Kern and Hainmueller 2006). Small parties with low office aspiration might therefore choose very negative

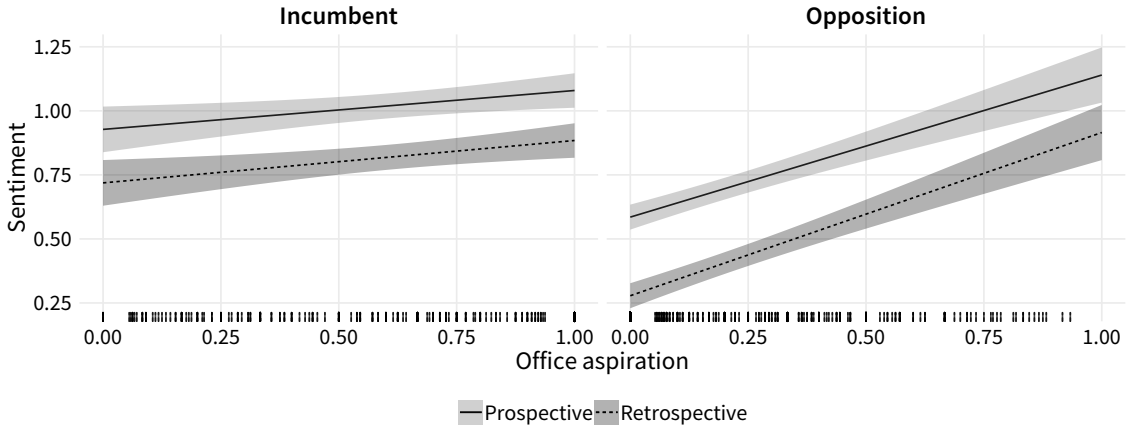
Figure 3.8: Predicting sentiment in prospective and retrospective manifesto sections, conditional on office ambition (national elections)



Note: Predicted values are based on Model 1 from Table 3.4. Shaded areas show 95 per cent confidence intervals.

campaign communication, outlining what will go wrong if the government (either the regional and/or on the federal incumbent) remains in office. Parties with low office aspiration on the subnational level might perceive such a 'doomsday' strategy as a promising way of maximising votes. On the national level, however, parties compete in very important first-order elections, and even the policy proposals of smaller parties receive considerable public attention. Thus, parties that have never/not often been in government in the past, might need to outline their goals for the future instead of explaining what will go wrong if the incumbent remains in office. Overall, the different impact of office aspiration on the national and regional level points to an important feature of multi-level governance. In first-order elections, parties need to provide promises for the future and act like they are 'serious' contenders to participate in future governments. In less important elections, however, parties without government ambition have higher incentives of 'going negative' and attacking incumbents' past, present, and future actions.

Figure 3.9: Predicting sentiment in prospective and retrospective manifesto sections, conditional on office ambition (German subnational elections)



Note: Predicted values are based on Model 2 from Table 3.4. Shaded areas show 95 per cent confidence intervals.

3.5.4 Robustness Tests

I conducted various robustness tests which mirror the findings above. First, I ascertain whether the choice of the dictionary influences the finding that opposition parties are more negative in retrospective sections, but not significantly more negative in prospective parts. Figures B.1–B.3 show the distribution of the *Sentiment* variable and estimate the correlations between both measures of sentiment on the level of prospective and retrospective manifesto sections. The correlations between the Lexicoder and the LIWC are strong and positive, ranging between 0.64 and 0.8 for method proposed by Proksch et al. (2019) and between 0.67 and 0.74 when employing Soroka's (2012) estimation of *Sentiment*. Moreover, I rerun Models 1 and 3 from Table 3.3 with the LIWC sentiment dictionary and the alternative aggregation of sentiment. The predicted values for the measurements of the dependent variables are similar across all models (Figures B.7 and B.8 which are based on the coefficients from Table B.8 and B.9). While the absolute values of *Sentiment* vary due to the different number of words (see Table 3.2) and the inclusion of 'neutral' terms, the relative differences between incumbents and opposition parties persist.

Table 3.4: Predicting sentiment in party manifestos, conditional on office aspiration

	M1 (National)	M2 (Subnational)
Opposition	-0.44^{***}	-0.34^{***}
	(0.09)	(0.05)
Class: Retrospective	-0.51^{***}	-0.21^{***}
	(0.08)	(0.03)
Office aspiration	-0.19	0.15^{*}
	(0.12)	(0.06)
Seat share	-0.00	0.01^{***}
	(0.00)	(0.00)
Year	0.08	-0.10^{**}
	(0.09)	(0.04)
Opposition × Retrospective	0.37^{*}	0.40^{***}
	(0.15)	(0.09)
Opposition × Office aspiration	0.28^{*}	0.01
	(0.11)	(0.05)
Opposition × Office aspiration × Retrospective	-0.34^{*}	0.07
	(0.14)	(0.07)
AIC	833.52	214.27
BIC	897.55	281.36
Log Likelihood	-403.76	-94.14
N	1018	1288
Num. groups: Manifesto	509	644
Num. groups: Election	130	141
Num. groups: Country	8	
Num. groups: Bundesland		16

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Note: Multilevel linear regressions with the sentiment as the dependent variable. Model 1 focuses on the manifestos from national elections. Model 2 reproduces the regression for German regional elections. Standard errors in parentheses.

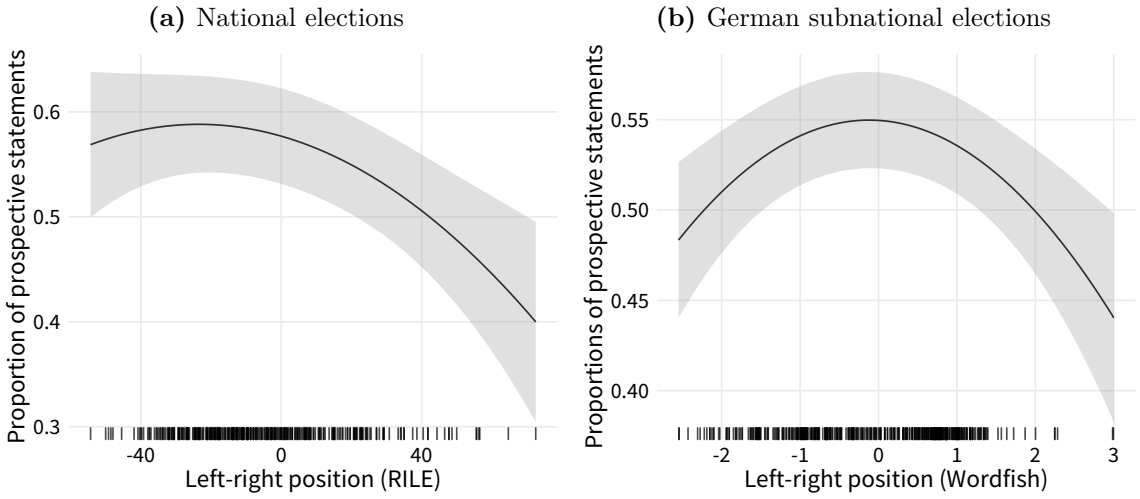
Recall that sentiment was aggregated to the level of classes in each manifesto. I also test whether the results change when I do not aggregate sentiment, but use the sentiment scores for each individual sentence as the dependent variable. Changing this model specification increases the number of observations from 1,138 (national) and 1,288 (regional elections) manifesto-class observations to 297,135 (national) and 576,187 (regional) sentence-level observations. Opposition parties are still significantly less positive in their entire manifesto, retrospective manifesto sections are more negative, and – importantly – the interaction term between *Opposition* and *Retrospective* remains large, negative, and statistically significant (Table B.10).

I also investigate whether certain democracies drive the results. Figure B.9 plots the predicted values from nine regression models – each regression reproduces Model 1 of Table 3.3, but excludes one of the countries before rerunning the model and plotting the predicted values. None of the countries seems to influence the aggregated results. Finally, I rerun the analysis with a differently conceptualised *Sentiment* variable. Instead of having two observations per manifesto – one for retrospective and one for prospective sentiment – I subtract each manifesto's prospective sentiment from the retrospective sentiment score. A higher value implies that sentiment in prospective sections becomes larger than sentiment in retrospective sections. If the assumptions of Hypotheses 1 and 2 hold, we should observe a larger value for opposition parties, as I expect the difference between both classes to increase for the opposition. Running two-sample t-tests and regression models offers support for this assumption (Table B.12). The coefficients of *Opposition* are positive and statistically significant across all models, indicating that differences between prospective and retrospective sentiment are larger for opposition than for incumbent parties.

An additional observable implication of the blame attribution hypothesis relates to the proportion of future-related statements by ideologically extreme opposition parties. Whereas moderate opposition parties might need and aim to outline their future promises in detail (Mansbridge 2003), extreme opposition parties could feel inclined to focus more on the past and present and the government's failure. As a consequence, the proportion of prospective statements should be smaller for extreme parties. To test this possible mechanism, I run multilevel linear regressions with the proportion of prospective statements as the dependent variable. I estimate separate models for the national and regional levels and only include opposition parties in the sample. Figure 3.10 plots the predicted proportions of prospective statements conditional on parties' left-right positions (RILE for national elections and Wordfish for subnational elections). In national and regional elections, we observe a curvilinear pattern with more extreme parties talking more about the past than

ideologically moderate parties. Ideologically right parties tend to talk significantly less about the future than parties located in the ideological centre. The difference of around 15 percentage points (national elections) and 9 percentage points (regional elections) corresponds to around 120 per cent of the standard deviations of the dependent variable (12 and 7 percentage points). Additionally, the positive, large, and statistically significant coefficients for *Seat share* show that larger opposition parties speak more about the future, even after controlling for the parties' ideology. Testing this additional observable implication supports the expectation that blame attribution strategies are less common across moderate and larger parties, and that parties with credible office aspiration must offer proposals for future policies.

Figure 3.10: The impact of left-right positions on emphasis of prospective statements



Note: Predicted values are based on Models 1 and 2 from Table B.11. Shaded areas show 95 per cent confidence intervals.

3.6 Discussion

Parties communicate their goals and positions during election campaigns. Media and voters pay attention to parties' campaign communication which is constituted, amongst other things, in party manifestos. When developing campaign strategies,

parties must make central decisions: is it more beneficial to focus on the past and present, or should a party rather outline policy goals in detail? In this paper, I ascertained how incumbency status and office aspiration influence these tactics.

Applying a new measure of prospective and retrospective statements to over 960,000 sentences from more than 1,200 national and regional party manifestos, I find support for the main theoretical expectations, both on the national and the regional level. First, party manifestos not only outline future policy, but contain large proportions of statements on the current situation and the past. Given that voters attribute blame and credit to previous actions and developments (Marsh and Tilley 2010), parties face incentives to highlight their own and their opponents' policy records. Considerable parts of manifestos are used for credit claiming purposes or to blame the government.

Second, incumbency status has a strong influence on parties' non-policy dimensions of election campaigns. Extending the studies by Crabtree et al. (2018) and Kosmidis et al. (2018), I show that opposition party manifestos are not consistently more negative than incumbent manifestos in their entirety but only in retrospective parts. Differences in sentiment become much smaller in prospective manifesto sections. In national elections, even opposition parties generally tend to frame the future positively. Only for the regional German elections do I find significant and substantial effects for office aspiration, suggesting that opposition parties with low chances or aspiration of becoming a government party only draft 'doomsday' manifestos in less important second-order elections. This finding speaks to the assumptions of promissory representation (Mansbridge 2003). Voters expect parties to outline pledges for the future. The more similar levels of positive sentiment in prospective manifesto parts suggests that even opposition parties with low office ambition in first-order elections face strong incentives to provide a positive outlook for the future and to outline their policy goals.

The results have broad implications for the understanding of party competition. Only recently have researchers shifted to manifesto analyses that go beyond saliency and policy positions in order to open up the 'black box' of intraparty politics (Crabtree et al. 2018; Eichorst and Lin 2019; Kosmidis et al. 2018). Yet, these studies could not distinguish between retrospective and prospective campaign communication. This study shows that parties aim to persuade voters and shift attention to their own achievements. Parties use retrospective and prospective statements in different ways to 'set the tone' for election campaigns. Often, voters have wrong perceptions of parties' positions (Dahlberg 2013; Spoon and Klüver 2017) and struggle to recall the fulfilment of pledges (Thomson 2011; Thomson and Brandenburg 2018; Naurin and Oscarsson 2017). I underscore that incumbents actively try to improve their reputation and justify past actions using credit claiming strategies in retrospective manifesto sections. Comparing the results from Proksch et al. (2019) with the findings in this paper reveals that differences in sentiment between incumbents and the opposition are much larger in political speech than in manifestos. The results from this paper can thus be regarded as rather conservative estimates that are likely to increase in different types of political statements. Future studies should analyse how retrospective and prospective rhetoric in parliamentary speech develops throughout the legislative cycle, and whether parties' press releases and televised leaders' debates are dominated by credit claiming and blame attribution. Moreover, conjoint or vignette experiments could reveal whether credit claiming or blame attribution strategies are more appealing to voters than descriptions of promises for the future. While we know that voters think in terms of blame and credit and that parties indeed mention these dimensions in their campaigns, we do not know whether voters appreciate that parties emphasise their own achievements or failures by political competitors. This paper points to the need that future research should focus more on non-policy dimensions of election campaigns and the influence of credit claiming and blame attribution on voter perceptions of parties.

Chapter 4

Media Coverage of Campaign Promises Throughout the Electoral Cycle

Abstract

A growing body of work shows that governments fulfil a majority of their promises outlined in party manifestos. However, only a minority of voters believes that politicians try to keep their promises, and many voters struggle to accurately recall the fulfilment or breaking of salient campaign pledges. I argue that this disparity between public perception and reality is partially driven by the information voters receive in the media. I also expect that news outlets focus more on broken than on fulfilled promises. Based on a new text corpus of over 480,000 sentences on pledges published in 25 newspapers during 33 electoral cycles in Australia, Canada, Ireland, and the United Kingdom, I find support for these expectations. The number of statements about broken promises is between 1.5 and 2 times as large as the number of statements about fulfilled promises, indicating a consistent negativity bias in all countries. Moreover, tabloid newspapers express much more negative sentiment in statements on promises when the endorsed party is not in government, whereas partisan broadsheet papers do not change their style of reporting. The results have important implications for studying negative information in mass media, election pledges, and the linkage between voters and parties.

4.1 Introduction

How do the media report on election promises throughout the electoral cycle? News outlets are one of the most important intermediary actors that establish a relationship between voters and parties (McCombs and Shaw 1972; Strömbäck 2005; Håkansson and Naurin 2016; Prior 2017; Soroka and Wlezien 2018). The media take a crucial role in providing information required to evaluate parties' achievements and failures (Soroka and Wlezien 2018). The way the media report on parties' ability to fulfil their campaign promises can influence citizens' assessment of government performance. The 'mandate model of democracy' and 'promissory representation' assume that voters retrospectively judge parties' performances during a legislative cycle by comparing campaign pledges with policy outputs. Citizens base their vote choice on parties' promises for the upcoming cycle, or a party's record of having fulfilled specific promises in the previous cycle (Downs 1957; Klingemann et al. 1994; Powell 2000; Mansbridge 2003; Grossback et al. 2005). This model of representation assumes that voters know about pledges for the upcoming legislative period and that voters are aware of fulfilled and broken promises. Sanctioning or rewarding parties requires that voters know about the most important problems and parties' record in parliament. Thus, voters need information to assess the performance of government.

The chain of delegation and accountability between principals (voters) and agents (representatives) requires a constant flow of information. Agents aim to inform the principal about their actions, and voters consult these signals to monitor agents. However, bias in reporting can distort the principal-agent relationship and undermine democracy. If voters do not receive the information necessary to evaluate the performance of parties, the normatively appealing theories of promissory representation and the chain of delegation might not work in reality. We therefore need to know how the media report on parties' promises and the ability of the government to fulfil these promises. This study focuses on three unanswered questions. First,

do news outlets focus more on fulfilled or broken pledges? Second, how does media coverage of election promises vary throughout and across electoral cycles? Third, does the style of reporting by partisan newspapers change when the endorsed party holds government office?

The role of the media could also explain the contradiction between scholarly findings on pledge fulfilment and the public perception of 'promise-breaking politicians'. Previous research concludes that parties indeed fulfil a large share of their manifesto promises (e.g., Thomson 2001; Praprotnik 2017; Klüver and Zubek 2018; Thomson et al. 2017). However, citizens usually do not read manifestos and do not know about the dozens of pledges stated in these official documents (e.g., Thomson 2011; Duval and Pétry 2018a). Instead, voters receive information on parties' actions from the media, not from parliamentary speech, legislative bills, or budget reports. Newspapers, television, radio, and online news inform citizens about policy outputs and political processes (Snyder and Strömberg 2010). As Kostadinova (2017: 637) notes, "media can help voters become more aware of the differences among political alternatives, thus facilitating the party-voter mandate linkage". A constant flow of information is necessary to hold governments accountable for policy actions. Despite the theoretical and practical importance of media for informing citizens about policy outputs, surprisingly, no study has analysed media coverage of election pledges throughout the electoral cycle.

Extracting and classifying over 486,000 sentences relating to political promises in 22 newspapers across 33 electoral cycles in Australia, Canada, Ireland, and the United Kingdom, I find robust evidence that media coverage peaks before general elections. Moreover, media outlets are significantly more likely to report on broken than fulfilled pledges. These results are stable across countries, newspapers, over time, and when considering only statements that explicitly mention a party or a politician. Finally, partisan tabloid newspapers express more positive sentiment when the endorsed party is in government. On the other hand, the composition of

the government does not seem to influence the way partisan broadsheets report on pledges.

The findings have important implications for understanding citizens' perceptions of election pledges and media coverage of political events. First, the finding that pledge coverage increases exponentially prior to elections is somewhat reassuring. Media indeed inform voters about promises, policy alternatives and 'what is at stake' during an election. Second, the very low degree of coverage on fulfilled promises might at least partially explain why the majority of citizens does not believe that parties keep their promises (ISSP Research Group 2018) and why citizens' evaluations of pledges are often inaccurate (Thomson 2011; Thomson and Brandenburg 2018; Naurin and Oscarsson 2017; Duval and Pétry 2018a). Third, the study speaks to the literature on negative and positive reports on political news. Previous research found negativity bias in reports about the economy (Soroka 2006), campaign coverage (Dunaway 2013), or local coverage of Presidential visits (Eshbaugh-Soha 2010). I add evidence to this body of work showing that the amount of coverage of broken promises significantly and substantially exceeds reports about fulfilled promises. This study offers a novel perspective on 'promissory representation' (Mansbridge 2003) and the mutual relationships between voters, parties, and the media.

4.2 Media and the Mandate Model of Democracy

In the classic version of the mandate model of democracy, voters choose between election programmes/platforms and select the party that comes closes to their policy preferences (Downs 1957; Klingemann et al. 1994; Mansbridge 2003). In order to make such an informed decision, voters need to recall what parties have achieved, and need to know at least a subset of the parties' campaign promises (Manin et al. 1999; Strömbäck 2005). Citizens judge parties' performances by (not) voting for a party or candidate at the upcoming election. This democratic chain of delegation

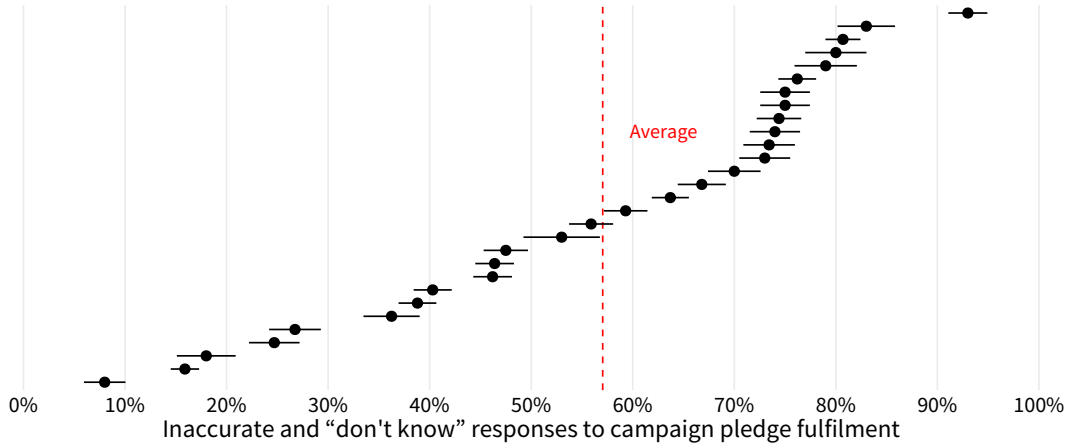
requires a flow of true information about parties' actions in parliament (Strøm 2000; Bovens et al. 2008). Without true and continuous information on the actions by the political agents, the principal cannot hold the agent accountable based on policy decision.

Previous studies assess the degree of mandate fulfilment based on election manifestos. The documents, published by almost all parties prior to elections, are the definite statement of parties' policy positions (Volkens et al. 2013) and provide a "well-defined and coherent body of officially sanctioned promises" (Pétry et al. 2018: 3). The documents are drafted in a lengthy process involving many intra-party actors, employ complicated language (Bischof and Senninger 2018), and contain dozens of election promises.[1] While manifestos are suitable to assess the degree of campaign pledge fulfilment, it is unrealistic to assume that voters remember specific manifesto content when voting retrospectively or prospectively. Evidence from a candidate study suggests that even politicians vastly underestimate the number of pledges in their own party's manifesto (Naurin 2016). Many voters struggle to accurately recall the fulfilment or breaking of salient pledges (Thomson 2011; Thomson and Brandenburg 2018; Naurin and Oscarsson 2017; Belchior 2018; Duval and Pétry 2018a). Based on the existing surveys, Figure 4.1 plots the percentage of inaccurate responses and the percentage of respondents who replied 'don't know' to the question as to whether a specific pledge has been fulfilled. We observe large variation in the degree of accuracy. For some pledges, less than 20 per cent of respondents made the wrong evaluation or replied that they do not know about the fulfilment. On the other extreme, some question about the fulfilment of specific pledges were answered inaccurately by over 80 per cent of respondents. Across all promises, 57 per cent of the respondents made an inaccurate evaluation or did not know whether a pledge has been fulfilled or not. On average, 27 per cent of the responses made an incorrect

[1] Reanalysing the comparative study of campaign pledges by Thomson et al. (2017), covering 204 manifestos, reveals an average of over 100 testable pledges per manifesto (with a range between 10 and 320 pledges). Figure C.1 plots the distributions for government and opposition parties.

evaluation. The mean proportion of 'don't know' responses across the statements amounts to 29 per cent. Citizens' knowledge of pledge fulfilment is limited.

Figure 4.1: The percentage of respondents who did not know about the fulfilment of a pledge or made an incorrect evaluation



Note: Each dot represents the proportion of responses that evaluted the fulfilment or breaking of a specific plege inaccurately or did not know about the fulfilment/breaking of the pledge. Data are derived from Thomson (2011), Thomson and Brandenburg (2018), Naurin and Oscarsson (2017), Belchior (2018), and Duval and Pétry (2018a). The content and proportions of inaccurate/don't know responses of each pledge are shown in Figure C.2. Vertical bars show 95 per cent confidence intervals. The 'partially' and 'fully fulfilled' categories are merged. Otherwise, the percentage for 'incorrect' evaluations would be even higher.

Rather than assessing the proportion of fulfilled promises, citizens are much more likely to judge parties' performances in government or opposition based on reports or comments received from intermediate actors, especially media outlets. The media play a crucial role in providing information about the fulfilment and breaking of pledges. As Prior (2017: 897) notes: "For a citizenry to be informed, information about government and elected officials must be available. Two critical sources of such information are news media and election campaigns." News coverage influences citizens' perception of political topics and problems (e.g., Chong and Druckman 2007; Walgrave et al. 2008; Ladd and Lenz 2009). Experimental evidence suggests that newspaper op-eds can have a long-lasting influence on public opinion (Coppock et al. 2018). Often, media set the agenda by establishing the salience of a policy

issue (McCombs and Shaw 1972; Boydstun 2013; Soroka and Wlezien 2018). Media also take a crucial gatekeeping function because journalists decide what decisions or opinions are newsworthy (Shoemaker and Vos 2009; Soroka 2012). Journalists function as partisan actors when deciding what (not) to cover in newspaper articles (Patterson and Donsbagh 1996). Analysing the method and extent of pledge coverage in the news helps us to understand why many citizens do not believe that parties keep their promises. If media mainly focus on unfulfilled or made, but not (yet) kept promises, the coverage could reinforce the common opinion that parties break campaign pledges.

4.3 Theoretical Argument

Having outlined the crucial role of media in the mandate vision of democracy, I next outline expectations and formulate testable hypotheses. The argument relates to different types of pledge coverage, the frequency of reports throughout the electoral cycle, and the partisan affiliation of newspapers.

I distinguish between three categories of coverage on election promises. The class 'fulfilled promises' includes sentences with a term indicating a promise and a word that indicates fulfilment. The class 'broken promises' contains a pledge-related word and a term that suggests that a promise was broken or not fulfilled. Statements about 'ongoing promises' mention a pledge made by a party or politician, provide updates about pledges made before an election, or broadly comment on a pledge. Importantly, sentences about ongoing promises do *not* include any information on the fulfilment or breaking of a pledge.

This study employs a broader conceptualisation than existing studies, and defines pledges as testable or subjective statements about future actions or outcomes.[2] I use the terms 'pledge' and 'promise' interchangeably throughout the paper. Broadening

[2]Thomson et al. (2017: 532) define a pledge as "a statement committing a party to one specific action or outcome that can be clearly determined to have occurred or not."

the definition for the analysis of newspaper coverage seems reasonable because voters tend to have a wider definition of pledges than scholars (Chapter 2). Citizens might perceive broad, non-testable media statements about promises as a pledge, which in turn influences their perception of parties and politicians.

4.3.1 Media Coverage of Broken, Fulfilled, and Ongoing Promises

The first expectation applies evidence on news reporting about party competition to the coverage of pledges. Electoral volatility and weaker party attachments require parties to respond to public opinion and to be sensitive about their issue positions. Parties are "challenged to provide clear-cut, distinguishing messages, in the hope that they grab media attention and make them stand out in public against their competitors" (Håkansson and Naurin 2016: 395). Through interviews, reports, direct or indirect contact with owners and interviewers politicians try to get 'their messages out' (De Beus 2011). Politicians aim to make comprehensible statements on a variety of policy issues and face incentives to state what they will change (or intend to change if they were in government). Parties act strategically and frequently outline plans and promises for the future, even if these goals are not very specific. Additionally, news outlets tend to frame politics as games or races with parties competing against each other (Patterson 1993; Hitt and Searles 2018). These 'game frames' occur both regarding the public approval of parties, but also regarding the policies offered to the electorate. Election promises are an effective means of reporting such races and the competition between parties. Therefore, I expect that reports on 'ongoing promises' constitute the largest share of pledge coverage.

H 1 (Ongoing Promises Hypothesis): *News outlets report more extensively about ongoing than about broken and fulfilled promises.*

The second expectation relates to the coverage of broken and fulfilled promises. Findings from psychology, economics, biology, and anthropology show consistent

evidence of negativity bias (for extensive reviews see Cacioppo and Garnder 1999; Baumeister et al. 2001; Rozin and Royzman 2001; Soroka and McAdams 2015). Humans react more strongly to negative than to positive information. This bias seems to result from journalistic norms of cynicism towards politics as well as consumers' preferences for negative news (Trussler and Soroka 2014). Negative framing of news has been observed in reports about the economy (Soroka 2006; Damstra and Boukes 2018), campaign coverage (Dunaway 2013), or local coverage of Presidential visits (Eshbaugh-Soha 2010). A similar trend should be observable when analysing pledge coverage. Especially opposition MPs point to the 'failures' of governments by highlighting promises that have been broken or that cannot be fulfilled until the end of the legislative cycle (Elmelund-Præstekær 2010; Thesen 2013).

Because of the competitive nature of liberal democracy, media act as watchdogs reporting on parties' decisions. From a normative perspective, journalists must report who is responsible for a policy output (Strömbäck 2005). Research has shown that news outlets tend to publish news that criticise political actors (Lengauer et al. 2011). For journalists, a broken pledge – even if the pledge is not highly salient – could be more 'newsworthy' than a fulfilled pledge of low importance. Thus, news outlets should be more likely to report about broken than fulfilled promises.

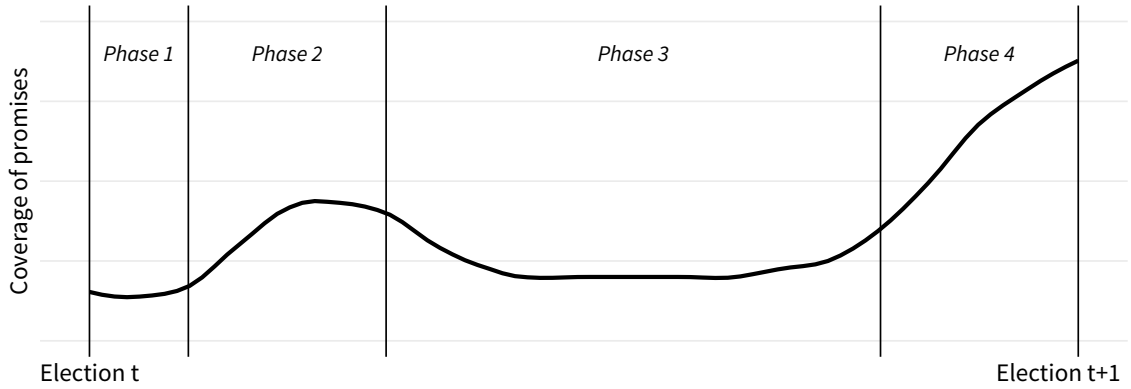
H 2 (Unfulfilled Promises Hypothesis): *News outlets report more extensively about unfulfilled election pledges than about fulfilled pledges.*

4.3.2 Media Coverage of Promises Throughout the Electoral Cycle

Besides differences in coverage on broken, fulfilled, and ongoing promises, I expect temporal variation in the amount of pledge coverage throughout the electoral cycle. I define the electoral cycle as the period between two general elections. Researchers have observed temporal patterns throughout the electoral cycle in several domains, such

as party support (Reif and Schmitt 1980; Müller and Louwerse 2018), partisanship (Michelitch and Utych 2018), or political communication strategies of coalition parties (Sagarzazu and Klüver 2017).

Figure 4.2: The expected coverage of promises throughout the electoral cycle



I expect that the reporting of pledge coverage is not constant throughout the cycle, but follows a curvilinear pattern with two peaks. Figure 4.2 plots a stylised version of the expected development of media coverage on promises. After *Election t*, media coverage on promises should be low (*Phase 1*). In single-party cabinets, important personall decisions must be made before policy-related decisions can be made. During this time, it is unclear what promises will be fulfilled, and media are more likely to comment, for example, on staffing decisions and the composition of the cabinet. In multi-party systems, before entering a coalition, parties negotiate with each other and regularly draft coalition agreements (Müller and Strøm 2008). During bargaining situations that do not allow for one of the preferred governments, coalition formation could last several months and require several attempts (Martin and Vanberg 2003; Ecker and Meyer 2017). In these cases, I expect that pledge coverage remains on a low level until the government has been formed. However, in the sample of countries selected in this study (see below) coalition negotiations usually did not take more than a few weeks. For this reason, I expect similar developments

of media coverage at the beginning of an electoral cycle for single- and multiparty cabinets.

Afterwards, parties should start fulfilling or breaking promises (*Phase 2*). The only existing study on the timing of pledge fulfilment (Duval and Pétry 2018b) provides anecdotal evidence that the largest share of pledges is fulfilled in the first year of an electoral cycle. As a result, coverage of promises should reach its first relative peak after around the first quarter of the cycle. Coverage on promises should remain on a stable, but lower level after this first peak in coverage (*Phase 3*). If a country holds mid-term or 'second-order' elections (either European Parliament or subnational elections) coverage should increase again. Yet, the level of reporting on pledges should be lower than before 'first-order' national elections.

The largest proportion of coverage is expected to occur in *Phase 4*, the time prior to *Election t+1*. Parties release manifestos in the months or weeks before the election (Dolezal et al. 2012). The launch of a manifesto serves the purpose of making the party's positions public. Parties hope that the public picks up some of the promises through the media (Däubler 2012: 58). News outlets compare parties' positions on salient issues, and summarise the promises by the major parties for the upcoming legislative period (Kostadinova 2017; Merz 2017). Additionally, politicians outline their plans for the future in interviews or press releases more frequently, and voters are interested in parties' pledges when an election comes closer (Bischof and Senninger 2018).[3] Besides the increase in coverage of promises for the upcoming election, politicians and journalists should also retrospectively judge the fulfilment or breaking of pledges in the current cycle. While government parties will claim credit for fulfilled promises, opposition parties will attribute blame and focus on unfulfilled pledges (Weaver 1986). Taking all of these factors together, coverage of ongoing, broken, and fulfilled promises should reach its peak before general elections.

[3]Bischof and Senninger (2018) show that online searches of manifesto PDF files peak in the weeks before elections.

H 3 (Electoral Cycle Hypothesis): *Media coverage about election promises increases until around the end of the first quarter and peaks towards the end of an electoral cycle.*

4.3.3 Newspaper Partisanship and Pledge Coverage

The ideological orientation of a newspaper could condition the coverage on promises if the endorsed party is (not) in government. Journalists might sympathise with one party, or editors expect journalists to write favourably about the endorsed party. Financial reasons incentivise partisan newspapers to follow the preferences of their readers (Strömbäck 2008). If a newspaper's endorsed party does not hold office, the outlet faces incentives to report more critically on a government's inability to keep their promises. Previous research shows that tabloid newspapers report more negatively about political issues and employ simpler language (e.g., De Vreese et al. 2006). Generally, tabloids should cover promises more negatively than broadsheets. However, I expect tabloids that endorse the government to apply more positive sentiment and focus more on fulfilled promises. Broadsheets, on the other hand, should usually be more balanced in their reporting of pledges, even when a broadsheet paper endorses the government. Therefore, I do not expect to observe a significant change in coverage when the endorsed party holds office.

H 4a (Tabloid Sentiment Hypothesis): *Tabloid newspapers employ more positive sentiment in statements on election promises when the endorsed party is in government.*

H 4b (Tabloid Pledges Hypothesis): *Tabloid newspapers put less emphasis on broken promises when the endorsed party is in government.*

4.4 Case Selection

Even in times of social media and online news, newspapers frequently set the agenda and provide comprehensive coverage of domestic and international political events (Roberts and McCombs 1994; Harder et al. 2017). Pledge coverage in print outlets is very likely to be replicated on the newspapers' websites and shared via their social media accounts.[4] Additionally, citizens still consult newspapers for information about politics. Prior to the 2016 Australian general election, over 50 per cent of respondents reported that they paid a good deal or some attention to newspaper reports about the election campaign. In the 2017 British Election Study, over 50 per cent of respondents stated that they regularly read about politics or current affairs in one or more newspapers. According to the 2015 Canadian Election Study, 70 per cent of respondents read newspapers at least once a week.[5] Newspapers permit the analysis of pledge coverage over many decades. Alternative sources, such as TV transcripts and social media, do not allow for comparable time-series analyses.

Recently, political scientists have applied methods from quantitative text analysis and natural language processing to massive corpora of newspaper articles in order to study phenomena such as the onset of conflicts (Mueller and Rauh 2018), the reporting on the economy or specific policy issues (Soroka 2006; Soroka and Wlezien 2018), the personalisation of politics (Langer and Sagarzazu 2018), the issue-definition stage of policy diffusion (Gilardi et al. 2018), or the estimation of political power (Ban et al. 2018). I follow this trend and analyse digitalised newspaper articles through quantitative text analysis.

I choose a comparative approach using outlets from Australia, Canada, the Republic of Ireland, and the United Kingdom. Although the analysis is limited

[4]The comparison of print and online articles of the same newspaper, retrieved through NexisLexis, confirms this expecatation. To avoid duplicates, the analysis therefore excludes online news from the selected outlets.

[5]Own calculations based on the respective election studies (Fournier et al. 2015; McAllister et al. 2016; Fieldhouse et al. 2018).

to English speaking parliamentary democracies, the countries offer considerable variation in important contextual variables about the democratic and media systems. The sample includes countries with a majoritarian constitutional design and strong executives (United Kingdom, Australia, Canada) and Ireland as a mixed multiparty system (Powell 2000). The electoral rules also differ across the cases: the United Kingdom and Canada employ a first-past-the-post (FPTP) system with single-member constituencies. Australia uses the Alternative Vote, elections in Ireland are conducted under the Single-Transferable Vote system.[6] As a consequence of the electoral system, the effective number of parliamentary parties and the number of government parties differ across the sample (Figures C.3 and C.4). During the period of investigation, Ireland was governed by coalition governments. Australia experienced both single- and multiparty cabinets. In Canada, only single-party cabinets were in power. Apart from the coalition between the Conservative Party the Liberal Democrats (2010–2015), all cabinets in the United Kingdom were single-party majority governments. The sample includes both broadsheet and tabloid newspapers that endorsed or disapproved the government. Despite a dominance of single-party cabinets in the United Kingdom and Canada, the chosen countries vary in terms of party-system and constitutional features, the degree of federalism, and government types.

I analyse newspapers from the United Kingdom to test Hypotheses 4a and 4b. For two reasons, the United Kingdom serves as an excellent case to test for differences between broadsheet and tabloids and to analyse whether an outlet's polarisation relates to the type of coverage. First, the majoritarian nature of the political system of the United Kingdom maximises 'clarity of responsibility' (Powell and Whitten 1993). Clarity of responsibility implies that citizens know who is responsible for

[6]Analysing New Zealand after 1996 could have beneficial in order to control for differences between personalised voting systems in the selected countries and a Mixed-Member PR system. Unfortunately, however, the availability of newspapers from New Zealand is much lower than for the other English-speaking parliamentary democracies.

broken and fulfilled pledges. Second, strong party newspapers have dominated the British newspaper landscape (Hallin and Mancini 2004). Research suggests that (changes in the) ideological positions of British news outlets have influenced voting decisions (Ladd and Lenz 2009; Reeves et al. 2016). In each election since 1992 at least three newspapers supported the incumbent, and at least two newspapers endorsed the opposition party (Figure C.16).[7] The data for eight daily newspapers and two Sunday editions covering the entire ideological spectrum allow for a comprehensive analysis of polarisation and newspaper formats.

4.5 Data, Classification, and Validation

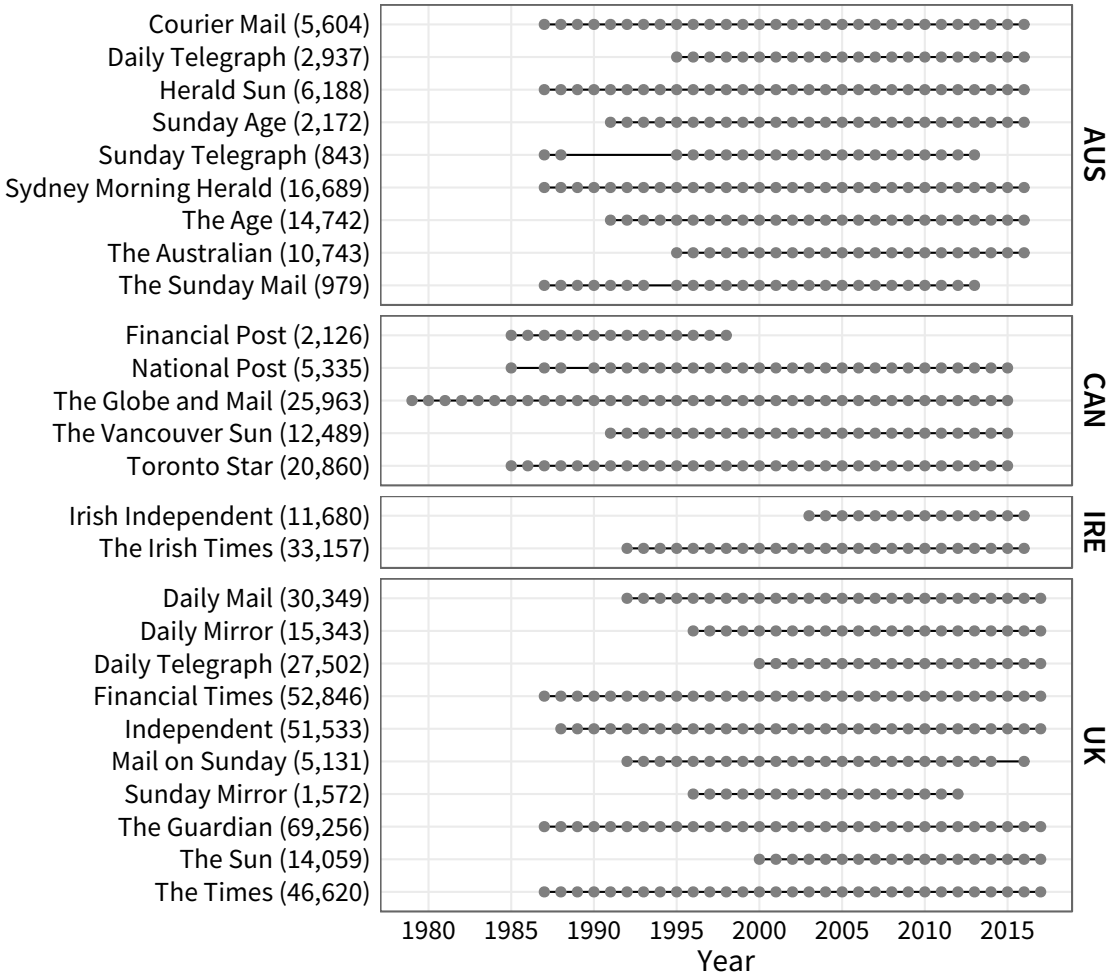
I determined the newspapers with the largest readership in each country and selected newspapers that are available on the online database NexisLexis. I downloaded all articles from these newspapers that contain a word indicating a pledge or promise, and that mention at least one of the parties represented in the national parliament of the respective country (for extensive information on the search query see Section C.2). Articles are retrieved throughout the entire cycle in order to map developments over time. Figure 4.3 plots the data availability for the 25 newspapers included in the analysis, along with the number of sentences on promises. The average newspaper availability amounts to 24.9 years (median: 25 years; standard deviation: 5.7 years).[8] With at least five newspapers per country, the coverage for Australia, Canada, and the United Kingdom is very comprehensive. The sample covers ten full cycles in Canada and Australia, seven cycles in the United Kingdom, and five electoral cycles in Ireland. The entire corpus includes 486,718 relevant sentences from 329,070 articles. On average, each article contains 1.5 sentences that directly relate to a promise.

[7] For the 2010–2015 electoral cycle, the United Kingdom was governed by a coalition between the Conservatives and the Liberal Democrats. For this cycle, I define the coalition as the incumbent. In all other cases, only one party governed.

[8] I treat the Sunday edition of newspapers as separate news outlets. If I merge weekday and Sunday editions, the mean and median remain unchanged.

News coverage remains remarkably stable within each country when dividing the number of articles by the number of newspapers and the length of a cycle (see Table 4.1).

Figure 4.3: Newspapers included in the analysis (number of promise-related sentences in parentheses)



Analysing almost half a million sentences mentioning promises and parties poses several challenges. A pure hand-coded approach is beyond scope of feasibility. An entirely unsupervised text-as-data approach on the level of articles, on the other hand, fails to pick up nuances and subtle differences between statements. I opt for a combination of human annotation and quantitative text analysis to filter relevant sentences from the corpus (Grimmer and Stewart 2013; Soroka and Wlezien 2018). Text pre-processing steps and the classification were conducted using the `quanteda`

Table 4.1: Descriptive statistics for each electoral cycle

Country	Start of cycle	End of cycle	Gov. Parties	Sentences	News outlets	Sentences (per day)	Sentences (day and outlet)
Australia	1987–07–11	1990–03–24	1	2,226	5	2.26	0.45
Australia	1990–03–24	1993–03–13	1	4,526	6	4.17	0.70
Australia	1993–03–13	1996–03–02	1	5,591	9	5.15	0.57
Australia	1996–03–02	1998–10–03	2	5,805	9	6.14	0.68
Australia	1998–10–03	2001–11–10	2	5,981	9	5.27	0.59
Australia	2001–11–10	2004–10–09	2	6,533	9	6.14	0.68
Australia	2004–10–09	2007–11–24	2	7,773	9	6.81	0.76
Australia	2007–11–24	2010–08–21	1	6,501	9	6.49	0.72
Australia	2010–08–21	2013–09–07	1	8,392	9	7.54	0.84
Australia	2013–09–07	2016–07–02	4	7,569	7	7.36	1.05
Canada	1979–05–22	1980–02–18	1	500	1	1.84	1.84
Canada	1980–02–18	1984–09–04	1	2,783	1	1.68	1.68
Canada	1984–09–04	1988–11–21	1	7,468	4	4.85	1.21
Canada	1988–11–21	1993–10–25	1	9,922	5	5.52	1.10
Canada	1993–10–25	1997–06–02	1	7,790	5	5.92	1.18
Canada	1997–06–02	2000–11–27	1	5,336	5	4.19	0.84
Canada	2000–11–27	2004–06–28	1	7,527	4	5.75	1.44
Canada	2004–06–28	2006–01–23	1	5,131	4	8.94	2.23
Canada	2006–01–23	2008–10–14	1	6,627	4	6.66	1.67
Canada	2008–10–14	2011–05–02	1	3,959	4	4.26	1.06
Canada	2011–05–02	2015–10–19	1	9,730	4	5.97	1.49
Ireland	1992–11–25	1997–06–06	2	6,422	1	3.88	3.88
Ireland	1997–06–06	2002–05–17	2	7,173	1	3.97	3.97
Ireland	2002–05–17	2007–05–24	2	10,750	2	5.86	2.93
Ireland	2007–05–24	2011–02–25	3	8,071	2	5.88	2.94
Ireland	2011–02–25	2016–02–26	2	12,421	2	6.80	3.40
UK	1987–06–11	1992–04–09	1	23,130	6	13.11	2.19
UK	1992–04–09	1997–05–01	1	38,271	8	20.71	2.59
UK	1997–05–01	2001–06–07	1	46,322	10	30.92	3.09
UK	2001–06–07	2005–05–05	1	44,109	10	30.89	3.09
UK	2005–05–05	2010–05–06	1	60,511	10	33.12	3.31
UK	2010–05–06	2015–05–07	2	63,304	10	34.65	3.46
UK	2015–05–07	2017–06–08	1	38,564	9	50.54	5.62

R package (Benoit et al. 2018). Having retrieved the newspaper articles, the relevant

content is extracted in several steps. First, all articles are imported into a text corpus. The corpus was reshaped to the level of sentences resulting in over 16.6 million sentences. After compounding names of parties and politicians, the corpus was transformed into a document-feature matrix. Since the main interest lies in sentences that mention promises, I applied a dictionary analysis and filtered statements that contain keywords indicating a pledge (the classification is described extensively in Section C.2).

Sentences relating to promises are divided up into one of three categories: broken pledge, fulfilled pledge, ongoing pledge. To classify statements, I construct a dictionary with terms indicating an election pledge as well as the fulfilment/breaking of the promise. A statement is coded as 'broken' if it contains one or more of the following terms: `broke, broken, fail*_to, failure, not_fulfil*, not_keep*, not_kept`. A statement that contains one of the following terms is coded as 'fulfilled': `fulfil*, kept, not_break*, not_broken`. I also chose an even more restrictive dictionary that only includes past forms of the verbs above. The differences between the 'broken' and 'fulfilled' classes become even stronger for each country (Figure C.9). Note that negations are rare. Non-negated terms occur around 10 times more often than the respective negations. Sentences which 'only' contain a term indicating a promises are coded as 'ongoing'. To sum up, all sentences include a term indicating a political promise, and I only consider articles that mention at least one domestic political party. In a second step, this subset of sentences is classified into three types of promise-related statements.

For two reasons, I selected only few keywords for the classification. First, all terms clearly contain information about the breaking or fulfilment when used in connection with the terms `promise*` or `pledge*`. A larger collection of keywords increases the proportion of false-positive classifications. Even the synonyms of 'to break' and 'to fulfil' often do not co-occur with promises in reports on politics. While unsupervised methods such as topic models (Blei et al. 2003; Roberts et al. 2014)

are more suitable for exploration and framing analysis, dictionaries are a reliable instrument to classify texts into known categories if the usage and meaning of words from the dictionary remain stable over time (Soroka and Wlezien 2018). Second, recent research shows that a small dictionary, limited to terms directly related to the concept can outperform larger collections of keywords. Muddiman et al. (2018: 10) note that "[l]arge dictionaries are not always necessary, but it is necessary that the content in the dictionary theoretically aligns with the concept being measured".[9] Without doubt, the fulfilment and breaking of promises is also expressed in different ways. Yet, the classification limited to key terms ensures that the sentences classified as broken or fulfilled relate to the concept under investigation.

The dictionary-based analysis resulted in a more reliable classification than supervised classifiers when comparing the results to 400 English crowd-coded sentences, randomly sampled from the text corpus (see SI Section C.3.2).[10] The average F1 score across the three classes amounts to 0.63. The class 'broken' has the highest F1 score (0.70), for class 'fulfilled' has the lowest F1 score (0.61). The considerably lower score for the 'fulfilled' class occurs because the dictionary classifies more statements as 'fulfilled' than human coders.[11] The automated classification somewhat overstates the 'true' number of sentences on fulfilled promises. As a consequence, observed differences in reports on broken and fulfilled promises should be even larger if all statements were handcoded.

In order to assess the face validity of the classification, I calculate the relative importance of terms in each class by calculating the term frequency-inverse document

[9]For instance, Muddiman and Stroud (2017) classify swearing in millions of tweets with only four terms, which results in better results than larger dictionaries.

[10]Supervised probabilistic classification methods trained through human codings of sentences (for instance Naïve Bayes or SVM) revealed worse classification accuracies than the dictionary approach.

[11]Using the aggregated human codings, the crowd coded 41 statements as 'fulfilled', while the dictionary identified 68 sentences as statements about a fulfilled promise.

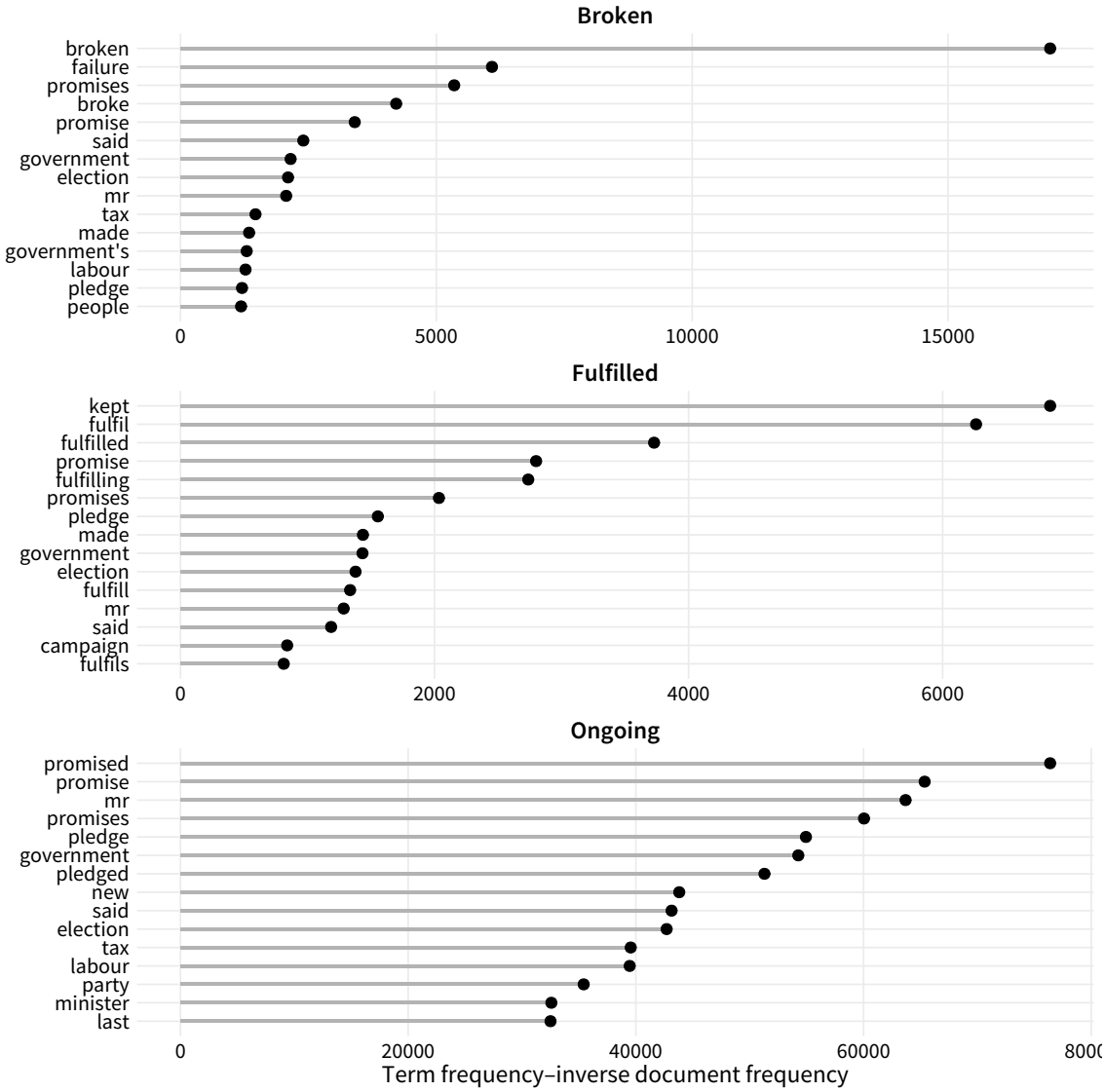
frequency (tf-idf).[12] A larger tf-idf implies a higher relative importance of a term in a class. Figure 4.4 provides an overview of the 15 most important terms in each class. Three out of the four highest tf-idf-weighted terms in the 'broken' category are `broken`, `failure`, and `broke`, whereas `kept`, `fulfil`, `fulfilled` and `fulfilling` are four of the highest tf-idf-weighted terms in the 'fulfilled' class. In the ongoing class, terms such as `promise*`, `mr`, `government`, `new`, `said`, `election`, `party`, or `minister` have very high tf-idf values. All of these terms either relate to a political actor or an action, which supports the assumption that most sentences in this class contain statements about future promises, or the progress in implementing policy.

Validation is essential when working with text-as-data approaches (Grimmer and Stewart 2013). I validate the selection and classification of articles in three ways (Section C.3). First of all, I test whether the chosen method of retrieving relevant newspapers picks up the relevant newspaper articles. For the three months before the 2015 UK General election I download all articles that mention keywords from at least one of the six pledges analysed in Thomson and Brandenburg (2018). Between 20 and 55 per cent of these articles also mention the terms `pledge` or `promise*` (Figure C.6). Thus, a large share of articles on these policy issues also included keywords used for retrieving the sample of newspaper articles. The search query based on promise-related seed words appears justifiable.

Second, I assess the potential shortcoming of selecting only a handful of keywords. When outlets report on fulfilled promises, they might simply describe the policy output without specifically mentioning that this policy has been a pledge. Articles on promises that have been classified as fulfilled in Thomson and Brandenburg (2018) should contain a higher ratio of words indicating fulfilment to words indicating the breaking of a pledge. If an article contains more words on fulfilled than on broken

[12] The tf-idf is calculated as the product of TF (term frequency) and IDF (inverse document frequency), where $TF = \frac{Frequency\ of\ term\ t\ in\ a\ document}{Frequency\ of\ terms\ in\ the\ document}$ and $IDF = log_{10} \times \frac{Number\ of\ documents}{Number\ of\ documents\ with\ term\ t}$.

Figure 4.4: Term-frequency inverse document frequency of the three pledge classes (15 terms with highest tf-idf values)



promises, I classify the text as covering the fulfilment of a promise. Articles that do not contain any term of the 'fulfilled' category, but one or more words on 'broken promises' are classified as 'broken'. As Figure C.7 shows, the focus on the breaking of a promise is highest for the two pledges that have indeed be broken, followed by the two promises that have been partially fulfilled. As expected, the fulfilled promises have the lowest ratios. This example provides anecdotal evidence that the classification into the three categories works meaningfully.

Third, I compare the automated classification of a random sample of 400 sentences to human codings collected through the crowdcoding platform Figure Eight (previously named CrowdFlower). The overall accuracy amounts to 0.65, and for none of the three classes does the balanced accuracy fall below 0.6 (Table C.2). While the measurement of the focus of a sentence is noisy, the classification error does not appear to systematically overestimate one of the categories. Due to this random misclassification, the differences in sentiment based on the automated classification should be rather conservative estimates, whereas the relative frequency of each class should reflect the actual occurrences in newspapers.

4.6 Variables and Modelling Approach

The first dependent variable is the *Focus on broken promises* which is estimated by comparing the frequencies of broken and fulfilled promises. I aggregate the statements for each cycle and newspaper by quarter, and divide the number of sentences about broken promises by the number of sentences about fulfilled promises. A value of 1 implies that a newspaper published the same number of sentences classified as fulfilled and broken promises. A value of 2 indicates that the number of sentences about broken promises was twice as large than the number of sentences about fulfilled promises.

In order to test how coverage varies throughout the cycle, I count the number of *Articles on promise-related sentences* per day for the three classes (ongoing, fulfilled, and broken). If no pledge-related sentence was published in any of the newspapers in a country, the count of sentences is set to 0 for this day. *Sentiment*, the third dependent variable, is measured with the Lexicoder Sentiment Dictionary (LSD) (Young and Soroka 2012). The LSD contains 2,858 word patterns indicating negative, and 1,709 word patterns indicating positive sentiment. Moreover, the dictionary includes 1,721 negated positive, and 2,860 negated negative terms. This

dictionary has been constructed to capture sentiment in news reports (Soroka 2012) and outperforms alternative sentiment dictionaries (Soroka et al. 2015; Proksch et al. 2019). I compound word patterns that indicate a positive/negative word preceded by a negation. Following Proksch et al. (2019) I measure the net positive sentiment as the logged ratio of positive to negative terms in a sentence (see also, Lowe et al. 2011):

$$Sentiment = log\left(\frac{\sum positive + \sum negations\ negative + 0.5}{\sum negative + \sum negations\ positive + 0.5}\right)$$

A positive value indicates that a sentence contains more positive than negative terms. If the number of positive terms equals the number of negative words (or if a sentence does not contain any words from the sentiment dictionary), *Sentiment* takes the value 0.

Electoral cycle, the main independent variable for Hypothesis 3, standardises the length of each cycle taking the value 0 for the day after an election and 1 for the day of the upcoming election. Rescaling the variable is necessary since the duration of cycles varies, either due to different maximum term lengths or early elections (Schleiter and Tavits 2016; Müller and Louwerse 2018). I add the second- and third-order polynomials of *Electoral cycle* to allow for curvilinear effects over time. In the models that estimate the ratio of reports on broken and fulfilled promises by quarter, I also add the *Average deviation of published polls* for government parties compared to the previous election result.[13] I consider the lagged quarterly *Change in GDP growth* as a proxy for recent economic developments. Economic downturns might correlate with more negative reporting on promises, or a decline of the economy might make it more difficult for parties to keep their promises (Thomson and Costello 2016). I

[13]The measure is based on the approach chosen by Müller and Louwerse (2018) using polling data from Jennings and Wlezien (2016, 2018). When a government party gained 30 per cent of the votes in *Election t* and the average standing throughout a quarter amounts to 25 per cent, the *Average deviation of published polls* equals 5. For coalition governments, I average the deviations for all coalition parties. Results remain unchanged when I estimate the percentage deviation from the election result or previous quarter.

classify the partisanship of each newspaper in the United Kingdom for each cycle using the data from Wring and Deacon (2010, 2018). Since the number of articles and extent of coverage varies across newspapers, cycles, and countries, I run multilevel regression (Bates et al. 2015) including random intercepts for countries, cycles, and newspapers.

4.7 Results

The results are presented in three steps. First, I provide descriptive evidence regarding the coverage of different types of reports on promises across the four countries. Second, I test the hypotheses outlined above. Third, I summarise the results from robustness checks and alternative variable specifications.

4.7.1 Coverage of Different Types of Promise-Related Statements

To what degree do newspapers cover broken, fulfilled, and ongoing pledges? Table 4.2 lists the number of relevant sentences for each country and category, along with the percentages of each class. Between 2.6 and 5.3 per cent of the statements include a term indicating a broken promise. The absolute number of sentences for this category ranges between 881 (Ireland) to 6,127 (United Kingdom). Statements about fulfilled promises constitute the smallest part, ranging between 1.6 and 2.4 per cent (see Table 4.2). Over 90 per cent of the statements that mention a promise and relate to domestic politics, do not contain any of the keywords used to code whether a sentence contains information on the breaking or fulfilment. Even though the 'ongoing' class only considers articles that mentioned a domestic political party and include `pledge*` or `promise*`, the number of sentences directly relating to political promises is certainly lower. However, decreasing the absolute frequency of statements by filtering based on additional keywords does not affect the relative proportions of

statements on 'broken' or 'fulfilled' promises.[14] The boxplots in Figure C.11 show the proportion of coverage for each class in one of the newspapers in a given cycle, again indicating that the relative coverage of each class is rather stable. Interestingly, however, newspapers in Australia – especially tabloids – are much more likely to mention the breaking of a promise, compared to outlets in other countries.

Table 4.2: Number of relevant sentences for each category and country

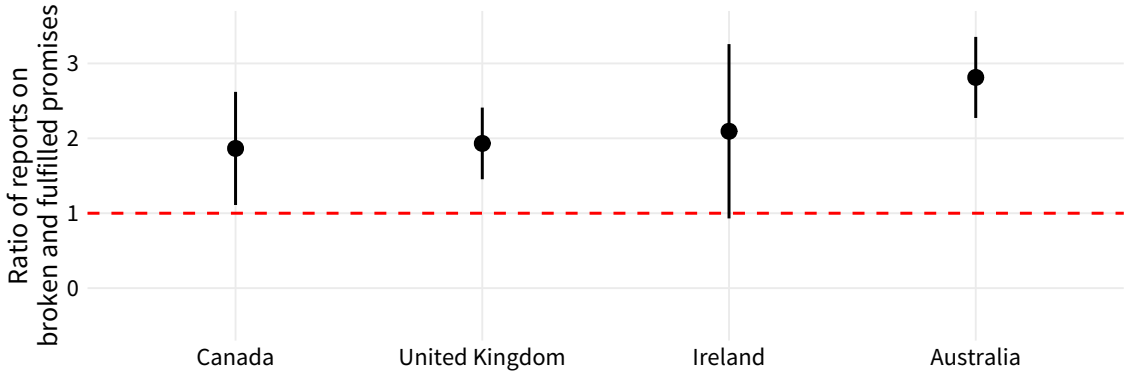
Country	Ongoing	Broken	Fulfilled	Total
Australia	56,710 (93.1%)	3,219 (5.3%)	968 (1.6%)	60,897
Canada	62,921 (94.2%)	2,261 (3.4%)	1,591 (2.4%)	66,773
Ireland	42,560 (94.9%)	1,396 (3.1%)	881 (2.0%)	44,837
United Kingdom	299,992 (95.5%)	8,092 (2.6%)	6,127 (1.9%)	314,211

Do newspapers focus more on broken than on fulfilled promises? To test Hypothesis 2, I run multilevel linear regressions with the ratio of statements about fulfilled and broken promises per quarter in each newspaper as the dependent variable. Without considering control variables, an intercept larger than 1 implies that coverage of broken promises is estimated to be higher than coverage of fulfilled promises. The intercept of Model 1 in Table 4.3 suggests that coverage of broken promises is over 2.2 times higher than coverage of fulfilled promises. The difference increases to 2.5 in Model 2, which removes newspaper-quarter observations with fewer than 50 promise-related sentences.

Model 3 adds dummies for each country and controls for the change in the economy compared to the previous quarter. Figure 4.5 plots the predicted values for each country, based on all the full sample of observations. The model estimates that newspapers in Canada, Ireland, and the United Kingdom, on average, include 2

[14] Introducing an additional filter, based on the verbs and adverbs with the highest keyness values (Bondi and Scott 2010) of 'prospective' rhetoric that also relate to promises (see Chapter 3), decreases the proportion of 'ongoing' promises to 86 per cent. Even when adding this filter, we observe that `pledge*` or `promise*` are mentioned much more often in connection with terms indicating political action than with terms indicating the fulfilment or breaking of promises.

Figure 4.5: Predicting the ratio of reports on broken and fulfilled promises per quarter for each country



Note: The predicted values are based on Model 3 of Table 4.3. A ratio of 1 implies that the number of statements on broken promises equals the number of statements on fulfilled promises. Vertical bars show 95 per cent confidence intervals.

times as many statements on broken than on fulfilled promises.[15] In Australia, the difference is even larger, indicating a stronger negativity bias.

When excluding newspaper-quarter ratios exceeding +4 and rerunning the model (limiting the sample to 67 per cent of the original dataset), coverage of broken promise still exceeds coverage of fulfilled promises by between 40 (Canada) and 80 (Australia) per cent (Figure C.13). The result do not depend on quarters that were dominated by extraordinary high coverage of broken promises. Considering the average poll change in a quarter for the government party or parties compared to the previous election (Model 4) reveals a negative and statistically significant coefficient. A one percentage point increase in government popularity, on average, decreases the ratio by –0.16 percent. If governments gain popularity, media tend put less emphasis on broken promises. Changes in the polls compared to the previous quarter is also negative, but not statistically significant with standard errors almost twice the size of the coefficient (Model 5).

[15]Looking at the descriptive statistics for each country (see also Figure C.12), the median ratio for these three countries exceed ranges between 1.14 (Canada) and 1.6 (Ireland). The mean ratios range between 1.98 and 2.08 (the median in Australia amounts to 2, the mean equals 3.02).

Table 4.3: Predicting the ratio of statements on broken and fulfilled promises

	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	2.20***	2.49***	2.75***	2.26***	2.11***
	(0.22)	(0.31)	(0.28)	(0.35)	(0.38)
GDP change (lag)			0.10	0.13**	0.13**
			(0.05)	(0.05)	(0.05)
Canada			−0.95*	−1.04*	−1.42**
			(0.47)	(0.47)	(0.51)
Ireland			−0.72	−0.40	−0.24
			(0.65)	(0.60)	(0.64)
United Kingdom			−0.88**	−1.32**	−1.47***
			(0.32)	(0.40)	(0.45)
Single-party cabinet				0.74*	1.10**
				(0.37)	(0.42)
Poll change cp. to prev. election				−0.16*	
				(0.07)	
Poll change cp. to prev. quarter					−0.06
					(0.11)
AIC	8491.34	6601.74	8580.65	6688.00	6035.40
BIC	8513.52	6622.84	8619.47	6741.44	6087.73
Log Likelihood	-4241.67	-3296.87	-4283.32	-3334.00	-3007.70
N	1892	1446	1892	1547	1385
Num. groups: Country-cycle	33	33		31	30
Num. groups: Newspaper	25	22	25	25	25

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Note: Model 1 only includes the intercept, Model 2 removes newspaper-quarter observations with fewer than 50 statements on promises. Model 3 adds GDP change and dummies for each country, Model 4 reproduces Model 3, but adds the type of government, and the average poll change of the government party/parties in the quarter compared to the previous election. Model 5 uses the poll change compared to the previous quarter. All models are multilevel linear regressions with random intercepts for country-cycles and newspapers. Standard errors in parentheses.

4.7.2 Coverage Throughout the Cycle

Does coverage of promises vary throughout the electoral cycle? I aggregate the results from each electoral period by standardising the duration of a cycle from 0 to 1. Figure 4.6 plots the development of pledge coverage throughout the cycle.[16] The black smoothed lines aggregate the cycles for each country. Although there is naturally some variation across the cycles (indicated by the grey lines), in almost all cases we observe a sharp increase of coverage prior to an upcoming general

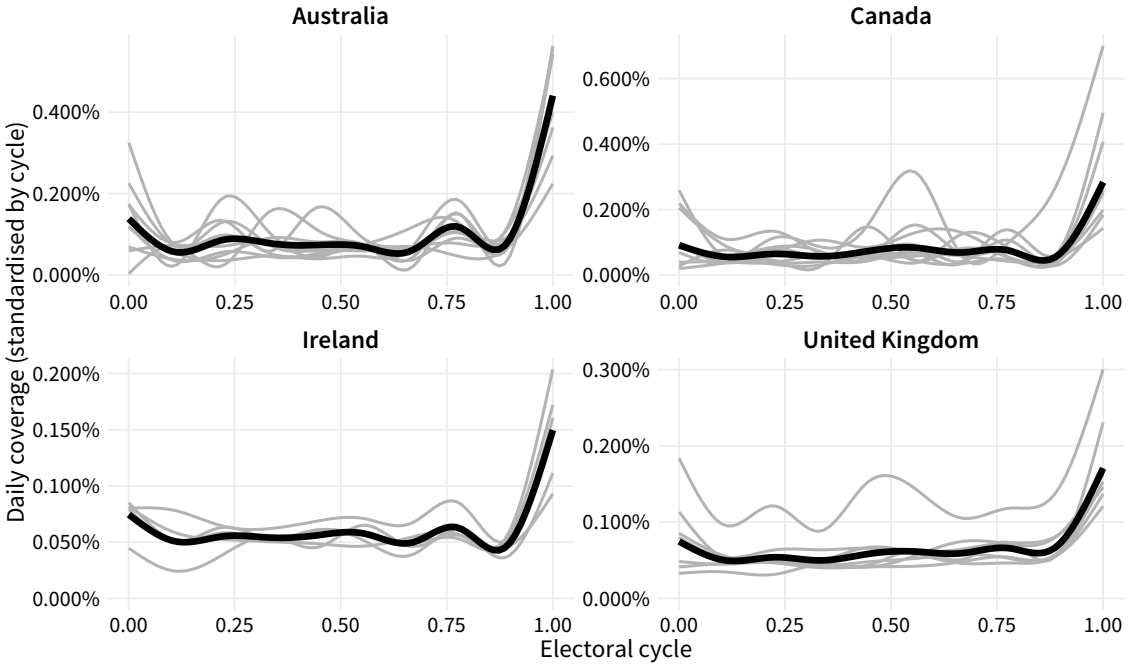
[16]I remove the Canadian electoral cycle from 1979 to 1980, since the cycle lasted only 272 days. The overall results do not change when including this cycle.

election. Figure C.10 plots the monthly count of articles for each country, and reveals very similar patterns. Prior to a general election, the coverage of election promises increases exponentially. Additionally, I run a set of negative binomial regressions that control for economic circumstances, the country, cycle, and months of cabinet reshuffles. Figure 4.7 shows the predicted values of pledge coverage based on the coefficients from Table C.4. The remaining continuous variables are held constant at their mean values, factor variables are fixed at their mode. Coverage of broken, fulfilled, and ongoing promises increases throughout the first quarter of the electoral cycle, decreases until the third quarter of the cycle and increases exponentially before general elections. The loess lines and negative binomial regressions thus offer only partial support for Hypothesis 3: while coverage peaks before elections, we do not observe a consistently higher degree of coverage at around the first quarter of a cycle, when parties (theoretically) should start fulfilling and break promises (Duval and Pétry 2018b).

4.7.3 Newspaper Endorsement of the Government

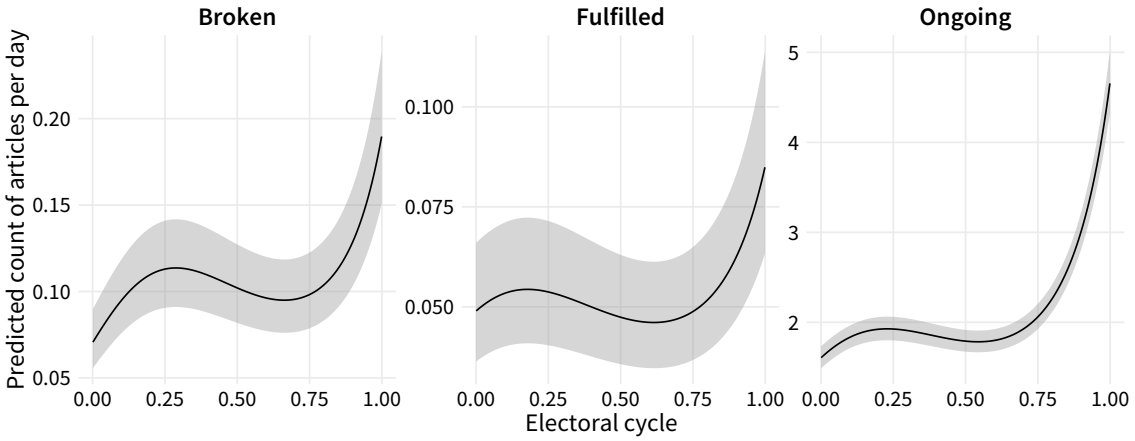
Does newspaper partisanship condition sentiment and the emphasis on broken promises? Using eight daily and two Sunday newspapers from the United Kingdom, I test the hypotheses regarding newspaper polarisation and party attachment. First, I analyse whether newspapers express more positive sentiment when the endorsed party is in office. The left-hand panel of Figure 4.8 plots the predicted values for sentiment based on the two-way interaction between the type of news outlet (broadsheet and tabloid) and the endorsement/disapproval of the government party (or the coalition that governed between 2010 and 2015). On average, broadsheet newspapers in the UK employ more positive sentiment than tabloids. As expected, tabloids employ considerably less positive sentiment when the endorsed government is in not power. When tabloids endorse the government, sentiment comes close to the tone by broadsheets. The estimated difference in sentiment for tabloids

Figure 4.6: The daily proportion of pledge-related statements throughout the electoral cycle



Note: The lines are based on generalised additive model (GAM) smoothers. Each line represents one electoral cycle. The black line shows the smoothed line aggregated for all cycles in a country. The dates are standardised from 0 to 1 for each cycle, where 0 marks the date of *Election t*, and the date of *Election t+1* is set to 1.

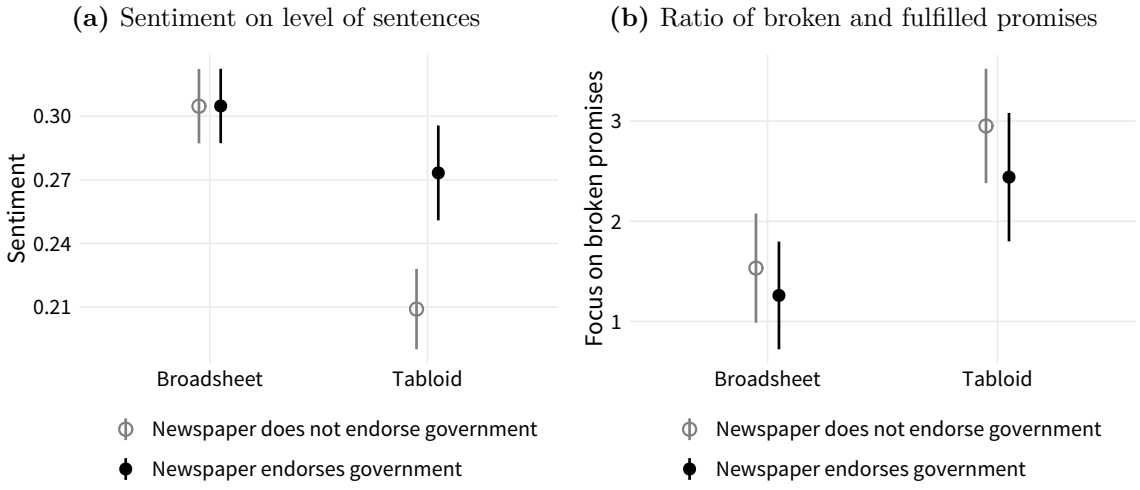
Figure 4.7: Predicting the number of articles per day, separately for each class



Note: The predicted values are based on negative binomial regression models with the number of daily articles as the dependent variable. Note that the y-axes differ for each facet to show the developments within each class. The figures are based on Models 1–3 of Table C.4. The grey areas show 95 per cent confidence intervals.

amounts to about 80 per cent of the dependent variable's standard deviation when aggregating sentiment on the newspaper-cycle level (0.077). The interaction is not only significantly, but also substantially significant.

Figure 4.8: Predicting the framing of coverage on promises in the UK

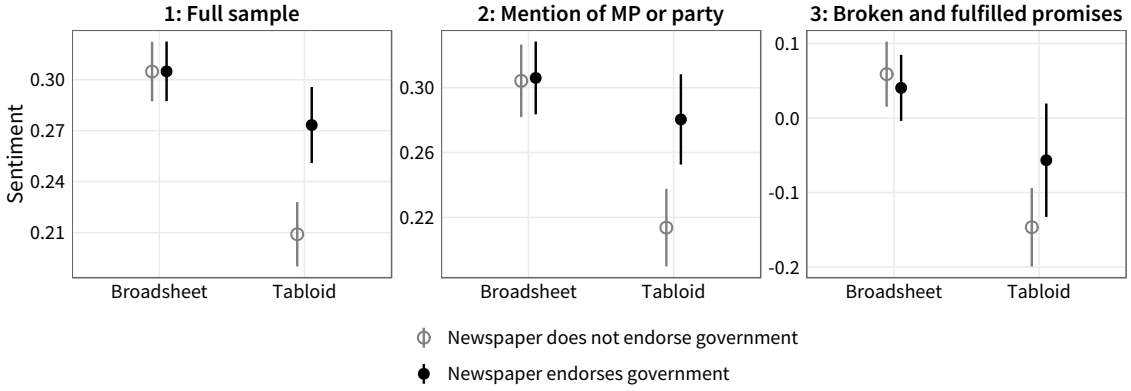


Note: The plots show the predicted values with 95 per cent confidence intervals which are based on Models 1 and 4 of Table 4.4.

In order to assess the robustness of this finding, I rerun the model for two subsets. Model 2 only uses statements (or its contextual unit of ± one sentence) that mention a British MP or party. Model 3 restricts the sample to sentences classified as 'broken' or 'fulfilled'. As the predicted values in Figure 4.9 underscore, the differences between broadsheets and tabloids remain, and the interaction term of *Newspaper endorsement* × *Tabloid* has the expected direction and is statistically significant in all three models (Table 4.4).

Turning from sentence-level sentiment to the quarterly ratio of fulfilled and broken promises, we observe a similar general relationship between the types of outlets (Figure 4.8). Tabloids generally focus more on broken promises than broadsheets. Yet, tabloids are not significantly and substantially more likely to focus on broken promises than broadsheets when the endorsed party is in government. The diverging findings between sentiment and the focus on broken promises could be explained

Figure 4.9: Predicting newspaper sentiment in promise-related statements for different subsamples



Note: The plots show the predicted values for sentiment with 95 per cent confidence intervals which are based on Models 1–3 of Table 4.4.

by the different approaches for measuring negativity bias. The conceptualisation of sentiment considers hundreds of positive and negative terms (Young and Soroka 2012), whereas the classification of broken and fulfilled promises rests on a handful of keywords. We should be more likely to observe shifts for large collections of positive and negative terms than for few keywords directly relating to promises. This finding points to the need to develop alternative classifications for the coverage of pledges in order to test the robustness of the dictionary-based method.

4.7.4 Robustness Tests

I conducted various robustness checks in order to test whether the results depend on the classification, measurement, or modelling choice (Sections C.5 and C.6). First, recall that all sentences included in the analysis contain a word indicating a pledge, and that the article mentions at least one domestic party. It might be possible that the pledge-related sentence does not immediately relate to a promise involving to a party or politician. I collected the names of all MPs serving in each parliament, and detected whether the pledge-related sentence or the contextual unit of ± one sentence mentions the name of a domestic party or an MP. Afterwards, I re-estimated

Table 4.4: Predicting sentiment and the relative emphasis on broken promises in the United Kingdom

	Model 1	Model 2	Model 3	Model 4
Intercept	0.30***	0.30***	0.06**	1.53***
	(0.01)	(0.01)	(0.02)	(0.28)
Newspaper endorses government	0.00	0.00	−0.02	−0.27
	(0.00)	(0.01)	(0.02)	(0.20)
Tabloid	−0.10***	−0.09***	−0.21***	1.42***
	(0.01)	(0.01)	(0.02)	(0.21)
Newspaper end. gov. × Tabloid	0.06***	0.06***	0.11*	−0.24
	(0.01)	(0.01)	(0.05)	(0.34)
AIC	898792.53	578075.15	40581.63	4067.11
BIC	898856.44	578136.40	40626.91	4095.95
Log likelihood	-449390.27	-289031.57	-20284.81	-2027.55
N	312086	200642	13993	904
N Groups: Cycle	7	7	7	7

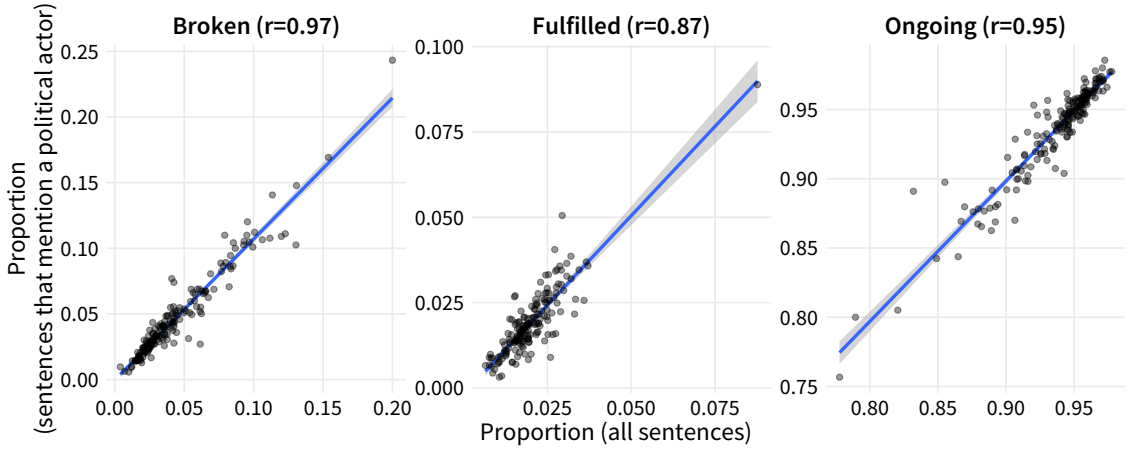
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Note: Model 1 predicts the sentiment for the sample of all promise-related sentences. Model 2 runs the model on the subset of statements that also mention an MP or a party from the UK. Model 3 only uses sentences that have been coded as fulfilled or broken. Model 4 predicts the ratio of sentences about broken and fulfilled promises, aggregated by quarters and newspapers. All models are multilevel linear regressions with random effects for each cycle. Standard errors in parentheses.

the proportions of coverage of each class. The correlation coefficients of aggregated proportions on the newspaper-cycle level amount to 0.95 for ongoing, 0.86 for fulfilled and 0.97 for broken promises (Figure 4.10). Plotting the predicted values for coverage over time for the sample of sentences that mentions a political actor does not change the conclusion. The shapes of the predicted count of articles are basically the same (Figure C.14).

To test for the robustness of pledge coverage throughout the cycle, I run separate regressions based on the type of government. The patterns reported above are similar for single- and multiparty cabinets (see Figure C.15). The only visual difference emerges for statements on 'broken' promises' in coalition governments. We see an increase after around 25 per cent of the electoral cycle, indicating that, on average, news outlets increase their coverage about broken promises. This could be driven by the observation that parties need to make compromises when entering government

Figure 4.10: Shares of promise-class across the three classes, depending on classification of relevant sentences



Note: The x-axis plots the proportions of the respective class when considering all promise-related sentences. The y-axis estimates the proportions based on sentences which *also* contain the name of national party or of a MP that served in the respective cycle (a window of ± sentence is considered to capture the context of a sentence). Each dot marks one cycle-newspaper observation.

which decrease their ability to fulfil manifesto promises (Thomson et al. 2017). However, since the majority of cabinets in the sample are single-party governments and since the loess regressions from Figure 4.6 mirror the predicted counts only at the end of a cycle, this trend should be taken with a grain of salt.

4.8 Discussion

How do the media cover election promises? The results of this study show that newspapers pay considerable attention to promises. Across 33 electoral cycles in four developed democracies, reporting on pledges reaches its maximum level of coverage before general elections. This is a reassuring finding in itself: newspapers indeed inform voters about parties' pledges made prior to elections, which is a crucial and necessary assumption of the mandate model of democracy (Mansbridge 2003; Håkansson and Naurin 2016; Kostadinova 2017). Only a small proportion of all statements contains information on a promise *and* its fulfilment/breaking, but media put more attention on statements about broken than on fulfilled promises. These

findings hold when I consider only the subset of actors that directly mention an MP or a party. Analysing newspaper polarisation in the United Kingdom, I find that tabloids generally apply more negative sentiment and focus more on broken promises. However, when the endorsed party is in government, tabloids frame statements on promises much more positively. Broadsheet newspapers, on the other hand, do not significantly change their style of reporting.

The results have important implications for the linkage between parties and voters. Håkansson and Naurin (2016: 395) assume that the "inherently moral aspect of promises also provides a foundation for media narratives of political success and/or failure, often featuring the typecast 'promise-breaking politician'". I find consistent support for the assumption that the media tend to stress political failures more than political success when covering promises. The extensive coverage of ongoing pledges might – at least partially – explain why so many voters do not believe that politicians keep their promises (ISSP Research Group 2018). Since the vast share of coverage is not devoted to broken or fulfilled pledges, voters might have the feeling that parties 'promise ever more' (Håkansson and Naurin 2016) without knowing about the actual performance of parties in office. The finding that most statements on promises do not contain information on the breaking or fulfilment helps us to understand why citizens' evaluations of pledges tend to be inaccurate (Duval and Pétry 2018a; Thomson and Brandenburg 2018). The media, one of the major sources for getting political information, rarely seem to state explicitly whether a promise has been fulfilled.

Although the selected countries show large variation in terms of the party system, federalism, and government types, the study is limited to a selection of English speaking parliamentary democracies. Future studies should analyse more countries that have experienced multiparty or minority governments in order to assess whether the countries are indeed generalisable. However, given the high similarity in terms of

coverage over time and between the types of statements, I expect that the finding should also hold in other democracies.

A second possible extension relates to the classification method. The dictionary approach chosen in this paper has the advantage that the categories of 'broken' and 'fulfilled' are clearly defined. The classification into categories depend on keywords that define the underlying concept (Muddiman et al. 2018). This advantage, however, is also one of the disadvantages: the choice of keywords could influence the results. I selected different sets of keywords and ran the analysis on different subsets of statements, resulting in consistent results. Nevertheless, more sophisticated deep learning methods that go beyond the bag-of-words approach and dictionary analyses (Mikolov et al. 2013; Goldberg 2016) should be applied to newspaper corpora to classify sentiment and attitudes expressed in reports on pledges.

Future research should also aim to analyse the framing of pledge coverage using topic models and supervised approaches (Blei et al. 2003; Roberts et al. 2014; Sagarzazu and Klüver 2017; Gilardi et al. 2018). Focusing on the content within each class could reveal differences in the statements labelled as 'ongoing promises' and uncover how specific promises evolve throughout a campaign. Coverage in newspapers could also be compared to the emphasis placed on issues in party manifestos to analyse whether salient pledges in manifestos also receive more media attention before an election. If congruence between saliency in manifestos and newspapers is low, we might have even more doubts as to whether the proportion of fulfilled manifesto pledges changes voters' perceptions of government performance. Finally, future research should combine newspaper coverage of promises with the literature on issue ownership (Petrocik 1996; Bélanger and Meguid 2008). If a party is perceived as being competent on a particular issue, the media might report even more often and more negatively on a specific promise in this policy area if the party fails to fulfil the pledge. This in turn could help to explain why voters do not believe that

parties keep their promises, and why government parties regularly lose public support throughout the electoral cycle.

Chapter 5

The Electoral Cycle Effect in Parliamentary Democracies[1]

Abstract

Does government party support decline in a monotonic fashion throughout the legislative cycle or do we observe a u-shaped 'electoral cycle effect'? Moving beyond the study of midterm election results, this is the first comparative study to assess the cyclical pulse of government party support in parliamentary democracies based on voting intention polls from 171 cycles in 22 countries. On average, government parties lose support during the first half of the electoral cycle, but at most partially recover from their initial losses. Under single-party government and when prime ministers control cabinet dissolution, support tends to follow the previously assumed u-shaped pattern more strongly. Finally, we find that government parties hardly recover from early losses since the 2000s.

[1]This chapter is co-authored with Tom Louwerse, and was published in *Political Science Research and Methods* (Müller and Louwerse 2018). Both authors have contributed equally to the work.

5.1 Introduction

Governance is a costly business in electoral terms. Existing studies show the average government party losing support at the next election (Nannestad and Paldam 1999; Wlezien 2017). At the same time, observers have found a u-shaped pattern of support for government parties throughout the electoral cycle: declining support at first, followed by a (partial) recovery (e.g., Tufte 1975; Campbell 1991; Erikson 1988). The strength of this 'recovery' is, however, a point of contention: does government support decline in a monotonic fashion (Mueller 1970; Hix and Marsh 2011; Döring 2016) or do we observe a u-shaped 'electoral cycle effect'?

Most work on this pattern of cyclical support for government parties in parliamentary democracies has analysed midterm or second-order elections, often in one or few countries (see, e.g., Reif and Schmitt 1980; Hix and Marsh 2007). Instead of looking at second-order election results to map party support during an electoral cycle, we capture parties' standing in national opinion polls. The unique cross-national focus is possible thanks to the availability of voting intention polls across many countries (Jennings and Wlezien 2016, 2018). This allows us to empirically examine the electoral cycle effect in much more detail than previous studies have done and to describe the conditions under which the electoral cycle effect is more pronounced.

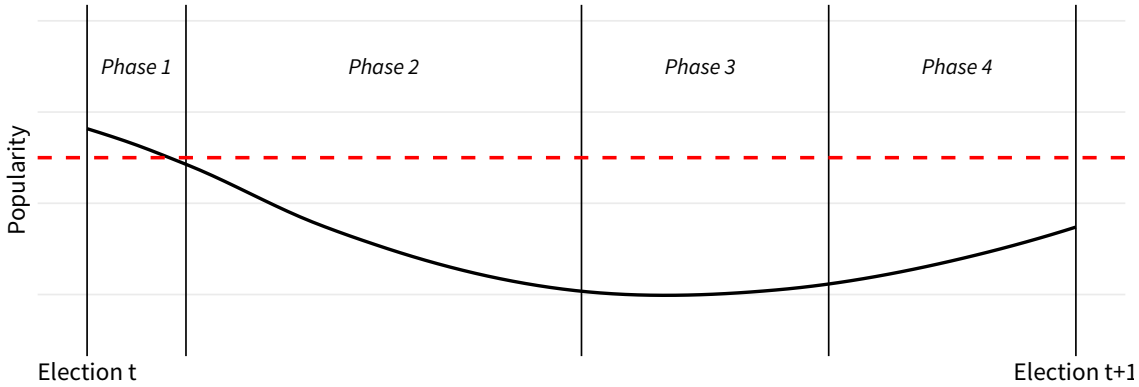
By analysing more than 25,000 opinion poll results for government parties from 171 electoral cycles between 1955 and 2017 we provide the first comparative study about patterns of government party support throughout the legislative cycle. We find that government parties tend to lose support during the first half of an electoral cycle, and can at most partially recover from these losses in the second part of the cycle. A small rise in support during the second half of the cycle tends to exist for single party governments and when the prime minister has the power to call early elections. Our analysis of general patterns as well as institutional and election-specific

determinants contributes to the literature on political representation, responsiveness, and the political business cycle.

5.2 The Electoral Cycle Effect

We define the expected pattern of support over a parliamentary term as the *electoral cycle effect*: government party support will decrease during the first part of an electoral cycle, and increase prior to the next election, while opposition party support will first increase and then decrease. Figure 5.1 plots a stylised version of what this cycle might look like under a single-party government. When a cabinet gets elected into office, the government is likely to enjoy a short 'honeymoon' period of public support for a few months (*Phase 1*). Afterwards, we expect support to start declining. Election campaigns may have raised hopes in the electorate, but the government might be unable to fulfil the expectations of many voters (Mueller 1970; Stimson 1976). Moreover, grievance-asymmetry theory suggests that voters punish government parties more strongly for things that go badly than rewarding them for favourable outcomes (Nannestad and Paldam 1999). As a result, the support for the government should decrease (*Phase 2*).

Figure 5.1: The expected popularity of government parties throughout an electoral cycle



Note: The electoral cycle ranges from the government inauguration date to the next election date.

Yet, we expect a partial recovery from these losses. Following a period of stable, but low support (*Phase 3*), government support should increase when the next election comes closer (*Phase 4*). Previous studies about US elections show that governments try to enact policies with highly visible benefits in the election year, while voters are likely to reward incumbents for high growth in the last six months, rather than the entire term (Healy and Malhotra 2013; Achen and Bartels 2016). Moreover, disapproval throughout a cycle may be a form of punishment by party supporters, but when the next election comes closer voters are more likely return to their former vote choice (Reif and Schmitt 1980; Miller and Mackie 1973). Increased partisan mobilization may also contribute to this recovery (Michelitch and Utych 2018).

We would not expect the electoral cycle effect to be equally strong under all circumstances. Based on the existing literature, we identify a number of moderator variables that we expect to affect the existence and strength of the electoral cycle effect. We argue, based on the 'clarity of responsibility' literature, that the electoral cycle effect is weaker if voters do not know which party to blame for government policy (Powell and Whitten 1993: 398). In majoritarian systems and under single party governments clarity of responsibility is higher than in coalition governments (Powell 2000; Lijphart 2012). It is more difficult for voters to determine which party should be held accountable for policies or the state of the economy when executive power is shared among more than one party. When a country is governed by a coalition we would expect the largest party to be identified most strongly with the government and therefore show a stronger electoral cycle effect than junior opposition parties. Lastly, we expect that a lack of clarity of responsibility will result in weaker electoral cycle effects under minority governments. These governments require support from at least part of the opposition to implement their legislative agenda, which makes it more difficult for voters to hold specific parties to account for government policies (De Vries et al. 2011).

Voters' abilities to assign responsibility for policy to parties is partly balanced out by parties' abilities to control the electoral cycle. Calling elections strategically can be advantageous for government parties: they can use the timing of the elections to maximise their support, for example by enacting policies that provide benefits in the election year or by holding elections when the economy is doing well (Schleiter and Tavits 2016). Although these attempts are not always successful, we should expect that on average higher party control over election timing results in a stronger recovery for government parties in the last phase of the electoral cycle. Therefore, we predict the electoral cycle effect to be more pronounced when the prime minister has strong institutional powers to decide on dissolution of parliament (Goplerud and Schleiter 2016). *All in all, we thus expect to see stronger electoral cycle effects under single party governments, for the largest parties in coalitions, for majority governments, and when the prime minister has strong dissolution powers.*

5.3 Data and Measurement

Although the precision of opinion polls for forecasting election results has been criticised, unquestionably polls play a role in (the media coverage of) politics, and within a margin of error show the relative standing of parties. We make use of recent datasets that contain poll data from legislative and presidential elections (Jennings and Wlezien 2016, 2018). We limit our analysis to legislative elections resulting in a sample of 22 countries, 171 electoral cycles and 25,333 polls for government parties (see Tables D.1 and D.2 for descriptive statistics). The unit of observation is a polling result for a government party. If more than one poll was published for the same party on the same day, the poll results are averaged. The dependent variable is the *Poll change* of each party calculated as difference between the current poll and the vote share at the previous election. For example, if party A gained 40% of the votes and drops to 30% in a poll, *Poll change* amounts to –10. The main

independent variable *Electoral cycle* measures the difference between the date when a poll was published and the date of the government inauguration divided by the actual length of the electoral cycle. The variable can thus take values between 0 (inauguration) and 1 (day of upcoming election). The expected curvilinear effect of the time in the electoral cycle on poll support is modelled by including the square of *Electoral cycle*. We also add the cubic value of the *Electoral cycle* to allow for more complicated patterns of government party support over the electoral cycle. Including the cubic value improves the model fit and mirrors the loess regression lines (Figure D.1) better.

We derive information on the type of government from the European Representative Democracy Data Archive (Andersson et al. 2014) and the ParlGov dataset (Döring and Manow 2018). The variable *Single party government* indicates whether the cabinet was formed by ministers from only one party. *Minority government* captures if a government commanded a majority in the lower house of the parliament. From the polling datasets (Jennings and Wlezien 2016, 2018) we recode a binary variable *Largest party* that indicates whether a party was the largest one in the coalition. The index developed by Goplerud and Schleiter (2016) operationalises the dissolution power of the prime minister (*PM dissolution power*) and the government's power to dissolve the parliament (*Government dissolution power*). Both indices range on a continuous scale between 0 and 10 with a higher number signalling stronger dissolution power. We also add *Party support at the last election* and the *Election year* to control for the possibility that the electoral cycle effect has changed over time and that the 'costs of ruling' (Stevenson 2002) increased in recent decades. The economic situation prior to an election has a considerable impact on vote choice (see, e.g., Lewis-Beck and Stegmaier 2000). An increase in GDP growth prior to an election is usually attributed to the incumbent resulting in increased popularity (Achen and Bartels 2016). We therefore merge each poll result with the economic situation during the lagged growth rate of the Gross Domestic Product (GDP)

compared to the previous quarter and add *GDP change* to all models. To check whether the relative GDP growth within a country leads to different conclusions, we also standardise *GDP change* by country and decade.

We test the relationship between the dependent and independent variables in a linear mixed-effects multilevel model with random intercepts for legislative cycles, parties and countries. As outlined above, our time variable (*Electoral cycle*) is expressed between 0 and 1 to allow for a comparison between election cycles and because countries and polls are taken at irregular intervals. Therefore, a panel data setup is not feasible and a multilevel model is the most appropriate choice. Autocorrelation of support for a party through a cycle is a potential issue that is explored in Section D.3 of the Supplementary Material: when we model autocorrelation using a Continuous Autoregressive (CAR1) structure, our results are comparable in substantive terms and the findings presented here remain statistically significant (Pinheiro and Bates 2000: 229). The basic model includes the following independent variables: *El. cycle*, *El. cycle*^{2}, *El. cycle*^{3}, *GDP Change*, *Party support at last election*, *Election year*. To test the expectations we add the interaction effects of the respective variable with *El. cycle*, *El. cycle*^{2} and *El. cycle*^{3}.

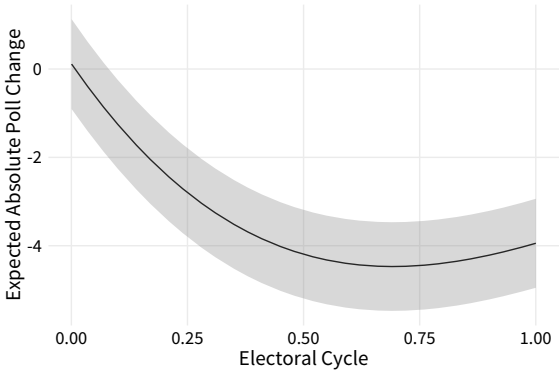
5.4 Results

Our basic model explains the change in support for government parties as a function of the electoral cycle. Figure 5.2a reports the expected values of support change across the electoral cycle, keeping all other variables at their mean or modes. The model predicts that on average government parties see a decline in party support over the first half of the electoral cycle. Afterwards, governments experience a very slight recovery, but cannot return to the levels when the government was inaugurated. The shape of the effect in the basic model mirrors the results of a bivariate analysis that fits a *loess* line to the relationship between the electoral cycle and government

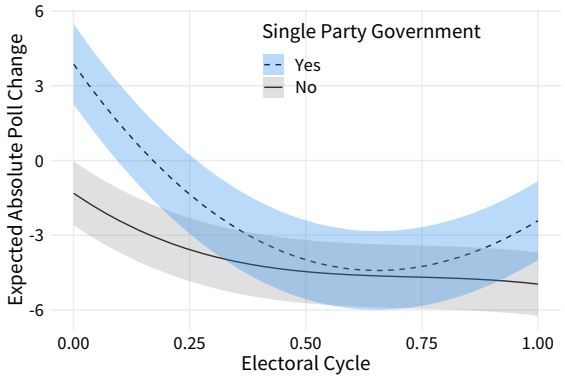
party support change (see Figure D.1). On average, we do not observe an increase in support prior to the next election. The similarity of the shape between the base model and the loess regression shows that the findings are not artefacts of our modelling strategy.

Figure 5.2: Moderated electoral cycle effects

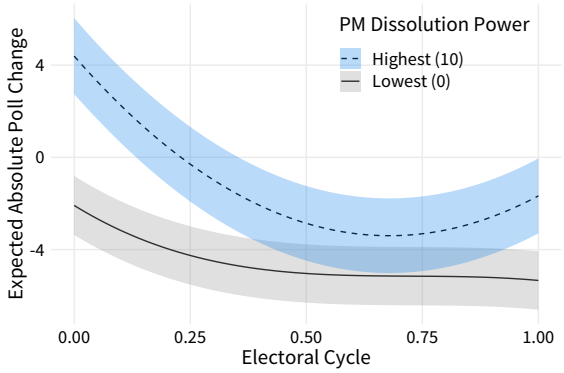
(a) Basic model



(b) Single party government



(c) Prime Minister dissolution power



Note: The bands around the lines show the 95% confidence intervals. Figures are based on Models 1-3 of Table D.3 in the Supplementary Material.

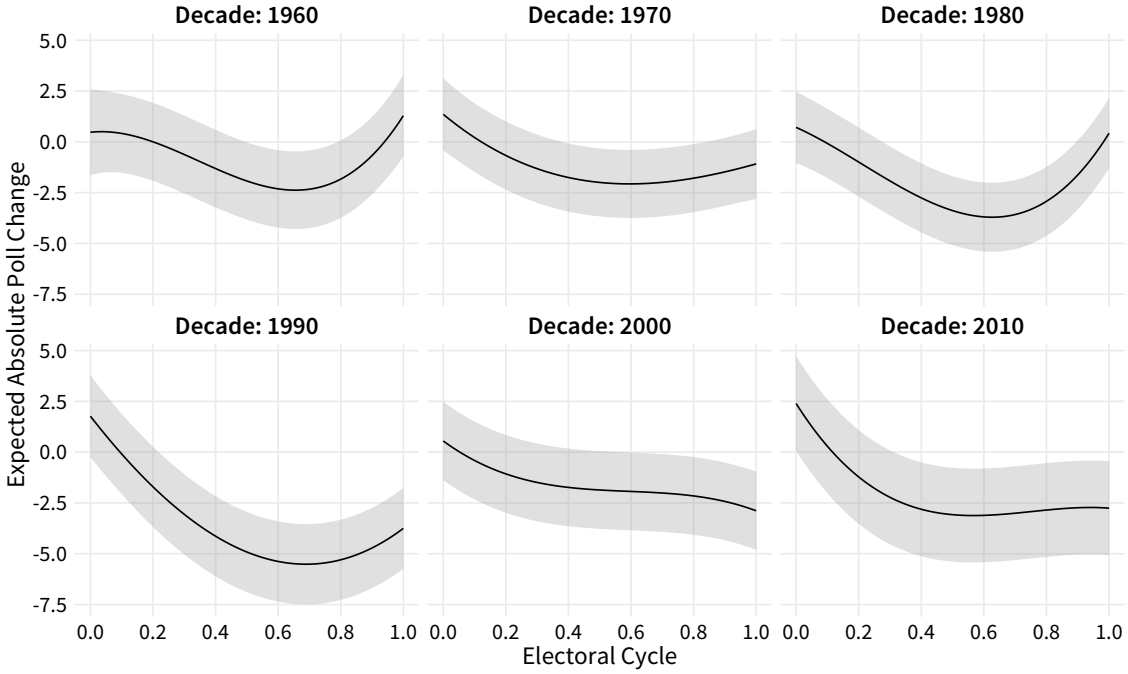
How does this electoral cycle effect differ between systems and governments? First, we compare single party and multi-party governments (Figure 5.2b). Single party governments show a somewhat u-shaped electoral cycle effect, starting off well with early gains, but gradually losing support over the electoral cycle. The lowest point is reached at around 70% of the electoral cycle, after which the single government party recovers somewhat. For other types of governments (coalition or

minority cabinets) we do not see a similar pattern. These government parties seem to be losing quite consistently over the electoral cycle, with only a slight, mostly linear decline. Against our expectations, neither being the largest party in the coalition nor minority government have a substantial impact on the electoral cycle effect (see Figures D.5a and D.5b).

Regarding the dissolution power of governments, our analysis points to a sizeable impact of prime ministerial dissolution power on the electoral cycle of government party support (Figure 5.2c). When prime ministerial control is at the highest level (10 on a 0–10 scale), we observe a u-shaped pattern of government party support, starting off at a high level of support. When the prime minister has no control over the election timing, there seems to be only a limited electoral cycle effect with some losses early in the cycle that level off towards the second half of the term. These effects are not observable when dissolution power lies with the government as a whole (Figure D.5e). The strategic use of election timing thus only seems to work when it is put in the hand of a single, unified actor.

Do we observe a change in the 'electoral cycle effect' over time? Having polling data for seven countries since the 1960s (Australia, Canada, Denmark, Germany, Netherlands, Norway, United Kingdom) allows us to test our model on this subsample and interact the absolute poll change with the decade of an election. Figure 5.3 shows that the u-shaped electoral cycle effect was stronger in the 1960s–1990s, and has almost disappeared entirely since the 2000s. The cost of governing increased without the prospect of earning back support, particularly in countries with single party governments and strong prime ministerial dissolution power. Thus, whereas in the past government support was more likely to follow a u-shaped pattern, a linear decline is prevalent more recently. One potential explanation is the decline of party identification (Dalton and Wattenberg 2000): dissatisfied government party supporters might have been more likely to 'return' to their party at the general election ballot box in the past than they are today. Another reason for this development is the

Figure 5.3: Historical development of the electoral cycle effect



Note: This model only includes countries with available polling data since the 1960s (Australia, Canada, Denmark, Germany, Netherlands, Norway, United Kingdom). Each panel includes all cycles that *ended* in the respective decade. The grey bands show the 95% confidence intervals. The Figure is based on Model 4 (Table D.4) in the Supplementary Material.

ideological convergence of political parties, especially in the United Kingdom. Parties' vote-seeking behaviour might lead them to converge 'too close' to the median voter, which reduces the clarity of the electoral choice between the two main contenders in a majoritarian system (Laver 2011: 494).

The Supplementary Material (Chapter D) reports several robustness checks; in all cases these confirm the results presented here. First, the patterns based on institutional clarity of responsibility are confirmed when conditioning the electoral cycle effect on Lijphart's (2012) executive-parties and federal-unitary dimensions. Second, the models presented above include one moderator variable for the electoral cycle at a time to avoid multicollinearity due to including multiple interaction effects. Adding all of the significant variables to a single model confirms the patterns reported above. Third, we performed an analysis that included the subset of 15 countries

for which we have at least three cycles. Fourth, we changed our measure of *Poll change* by looking at the percentage deviation from the election result, and ran an additional model with a lagged dependent variable. Fifth, instead of using the length of the actual cycle, we also measure the electoral cycle by using the 'planned' length (i.e., the maximum number of years of a legislative period) as the end point of a cycle. Finally, in many countries, the government can strategically set the election date which could influence the shape of the support throughout the cycle. We distinguish between regular and early (opportunistic or failure) elections (Schleiter and Tavits 2016). Regular elections show a modest u-shaped electoral cycle effect with losses early in the term and a slight recovery from about midway through the cycle. In opportunistically called elections we observe a recovery at the end of the cycle, whereas failure elections are characterised by a steady, almost linear, decline of government party support.

5.5 Discussion

Our results document the 'electoral cycle effect' throughout more than 170 legislative cycles. The often assumed u-shaped electoral cycle pattern predicted from the study of midterm and second-order elections holds only under certain circumstances. While we found consistent evidence that government party support decreases during the first part of an electoral cycle, parties recover from these losses at most partially. Our general pattern matches best with recent analyses of national government parties' results in European Parliament elections (Hix and Marsh 2011; Döring 2016). A curvilinear effect is more evident under single party governments and if the prime minister has high dissolution power. Small coalition partners tend to lose more support compared to the largest coalition party. Despite these general trends we still observe large deviations across countries and cycles. We hope that future research examines potential reasons for this variation more closely.

These findings are relevant to the broader comparative debate on the representative link between voters and party governments. Parties increasingly seem to have to choose between responsible and representative party government or, to put it slightly differently, between office and votes (Mair 2009). Whereas parties might accept that there is some cost to governing, the knowledge that it will be difficult to regain early losses might be particularly worrisome from the perspective of responsible party government. Further research is warranted to study the causes of these changes in the prevalence of the electoral cycle effect: is this an indirect effect of modern, critical voters, or do supply-side factors explain why government parties cannot recover their electoral losses?

Chapter 6

Conclusion

Classic models of political representation assume that retrospective assessments of party performance and evaluations of promises for the upcoming legislative cycle influence vote choice (Downs 1957; Manin et al. 1999; Powell 2000; Mansbridge 2003). Although this vision of democracy has normative appeal, voters seem to have only limited knowledge of parties' promises and employ other heuristics when casting their ballot (Thomson 2011; Thomson and Brandenburg 2018; Naurin and Oscarsson 2017). Moreover, while studies show that parties fulfil a large proportion of pledges (Thomson et al. 2017), most voters do not believe in the willingness or ability of parties to fulfil promises (ISSP Research Group 2008, 2018; Naurin 2011). Against the background of the contradictions between theoretical assumptions, empirical findings on pledge fulfilment, and survey evidence, this dissertation examined four research questions:

1. Do citizens and experts have different understandings of the concept of election pledges?

2. Under what circumstances do parties focus on future promises or rather engage in credit claiming and blame attribution?

3. Does the coverage and framing of election promises in the media exhibit a negativity bias?

4. At which periods throughout the electoral cycle does government support increase or decrease?

6.1 Findings

This dissertation chose a comparative approach analysing multiple countries and levels of governance, and applied a combination of quantitative text analysis (Grimmer and Stewart 2013), crowd-sourced coding (Benoit et al. 2016), and opinion polls (Jennings and Wlezien 2016, 2018) in order to answer central questions relating to accountability and the democratic mandate.

Having received access to the aggregated results from the expert reliability test, Chapter 2 uncovered differences in coding of sentences between experts and crowd workers. The results revealed disagreement both within and between pledge scholars and non-experts. Crowd workers have a broader understanding of promises, although they received the same coding instructions as experts and were monitored continuously in order to remove workers who coded statements randomly (Benoit et al. 2016). The findings not only show that pledge scholars and non-experts perceive pledges differently, but the results might also explain why the proportion of fulfilled promises is higher than assumed by the public. Many vague and non-testable pledges are perceived as a pledge by crowd coders, but are not considered as a pledge in previous analyses.

In Chapter 3, I showed that large parts of official campaign communication in party manifestos do not relate to the future. Instead, parties seem to engage both in prospective and retrospective communication, in order to highlight achievements, blame the government, and outline the future. Parties devote substantial emphasis to a description and evaluation of the past and present. On average, retrospective

statements – often consisting of credit claiming and blame attribution – make up half of the manifesto. Additionally, I qualify Crabtree et al.'s (2018) conclusion that incumbents employ more positive language. Differences in sentiment mainly occur in 'retrospective' parts of manifestos that governments use for claiming credit. Incumbents and opposition parties employ more similar levels of positive sentiment when outlining the future. The findings question a crucial assumption of the saliency theory (Budge and Farlie 1977): parties do not always 'talk past each other', but often engage in confrontation by highlighting past achievements or the government's failures. The study also speaks to the emerging literature on non-policy dimensions of campaign communication and the drafting of party manifestos (e.g., Dolezal et al. 2018; Harmel 2018; Eichorst and Lin 2019). Parties strategically make statements about credit claiming and blame attribution to persuade voters.

Chapter 4 provided the first comparative study of media coverage of political promises. A massive corpus of news articles on promises, classified through quantitative text analysis (Grimmer and Stewart 2013), uncovered a consistent negativity bias in reports on pledges. Focusing on newspaper polarisation in the United Kingdom, I also found support for the hypothesis that partisan tabloids (but not broadsheet newspapers) report more negatively about promises when the endorsed party is not in government.

Chapter 5 mapped party support throughout the electoral cycle by analysing parties' standing in national opinion polls. The analysis of over 25,000 polls revealed that, on average, governments lose most support in the first half of the legislative cycle. The assumed u-shaped development is more likely to occur when clarity of responsibility is higher. Comparing countries with available polling data since the 1960s reveals that since the 2000s, parties rarely recover from early losses in the cycle. More recently, governments tend to suffer electorally and are not likely to recover from early losses.

6.2 Implications

These findings have clear theoretical and empirical implications. First of all, the broader understanding of the concept of pledges by non-expert coders (Chapter 2) helps to explain the difference between empirical evidence of pledge fulfilment and public opinion. The concept of pledges obviously differs between instructed coders and scholars. Therefore, the proportion of fulfilled pledges, found in academic studies, might not be very telling. We require a better understanding of which pledges received public attention. The chapter also has important methodological implications. Crowd coding can serve as a means to reassess a concept and to refine coding instructions. Additionally, using the crowd to code promises, researchers can collect multiple codings in a fast and affordable way. As a result, we can add measures of uncertainty (Benoit et al. 2009b; Mikhaylov et al. 2012) and detect ambiguous promises that should be treated separately in empirical analyses.

The second implication stems from the classification of prospective and retrospective rhetoric (Chapter 3). Not all positional statements in manifestos are future-related statements. From a scholarly perspective, it is therefore highly problematic if blame allocation (retrospective and negative) or credit taking (retrospective and positive) for either alleged or actual achievements of parties is misinterpreted as the future policy position or even the pledge of a party. Researchers have mostly assumed that parties compete on issues and positions, but not necessarily on subjective assessments of the past and present. Yet, the findings imply that parties deliberately decide to attribute blame and claim credit, even in official campaign communication. Party competition goes beyond saliency and position taking (Dolezal et al. 2018), with retrospective and prospective rhetoric being a crucial dimension of campaign strategies. Political text in general, and party manifestos in particular, should not be treated as a collection of future-related statements or even promises.

Third, the observed negativity bias by media outlets (Chapter 4) directly speaks to the discussion about promissory representation. The fact that many voters struggle to recall the fulfilment of salient pledges "raises questions about the quality of the coverage of electoral pledges" (Duval and Pétry 2018a: 9). The observed trend of negative reporting on promises might explain the public perception of the 'pledge-breaking politician'. The results extend the conclusion drawn from the crowd-sourced concept reassessment. Citizens have a broader understanding of pledges. Additionally, the sources voters consult to assess government performance may not cover the types of promises analysed by pledge scholars. The media seem to serve as a gatekeeper and rarely cover the fulfilment or breaking of less salient promises.

The fourth implication relates to the results that governments lose most support at the beginning of the cycle and – in recent times – struggle even more to regain support prior to the upcoming elections (Chapter 5). The classic literature on midterm elections assumed a curvilinear shape of support (Tufte 1975; Erikson 1988; Campbell 1991). Yet, the finding that parties in parliamentary democracies struggle to recover from early losses has important consequences for representation. While anecdotal evidence shows that parties fulfil the majority of promises at the beginning of a cycle (Duval and Pétry 2018b), and although pork-barrel politics tends to increase prior to elections (Shepsle et al. 2009; Kang 2015), citizens often do not reward incumbents for these efforts. Parties losing support, irrespective of their performance in office, challenges the assumptions of promissory representation. Therefore, the findings speak to the debate on the link between voters and governments. Parties seem to have to choose between gaining office after an election or losing votes at the next election.

6.3 Future Research

The dissertation's innovative empirical approach and central findings allow future studies to tackle substantive research questions about representation. Future research should tackle four areas: temporality of political rhetoric; methodological advances in text classification; the analysis of pledges beyond party manifestos; and citizens' evaluations of prospective and retrospective rhetoric.

6.3.1 Temporality of Political Statements

First of all, more research should focus on the question of when parties engage in blame attribution and credit claiming. Shifting the focus from manifestos to political speech and leaders' debates will result in a better understanding of the strategic usage of prospective and retrospective rhetoric. While manifestos are official documents with relatively low degrees of negativity, differences in sentiment between incumbents and opposition parties is even higher in parliamentary speech (Proksch et al. 2019). Moreover, manifestos are drafted in a long process and few sentences are based on spontaneous decisions (Volkens et al. 2013). Political speech is structured differently. As long as an MP receives speaking time, politicians are free to express their personal views (Proksch and Slapin 2015). Whereas party manifestos are published only once every few years, speeches are given throughout the entire cycle and can reveal important temporal patterns (Bäck et al. 2019). For the theory of credit claiming and blame avoidance it will be highly relevant to understand whether and at what time incumbents start focusing on retrospective assessments, and whether blame attribution follows curvilinear patterns. Since politicians hope that their speech receives public attention, future studies should analyse whether prospective or retrospective statements increase the probability that a speech is covered in the news. Televised leaders' debates should also be analysed in terms of the temporal dimension. TV debates attract massive numbers of voters and usually

are amongst the most important campaign events (Schroeder 2008). While previous research has analysed negativity, incivility, and policy- versus person-focus (Brooks and Geer 2007; Maier and Jansen 2017; Maier and Renner 2018; Hopmann et al. 2018), we do not know when candidates decide to offer future policies rather than attacking their competitors. The field of analysing temporality in political statements is wide open.

6.3.2 Voters' Perceptions of Pledges and Retrospective Rhetoric

While Chapter 3 uncovered retrospective and prospective campaigning approaches, we do not know whether voters prefer to receive information on past achievements, failures, or promises for the future. Experimental evidence can capture the approval of different forms of temporal communication. For instance, conjoint experiments can be used to identify the types of temporal statements that are most appealing to voters (e.g., Hainmueller et al. 2015; Horiuchi et al. 2018). Given that the crowd workers received the instructions developed by pledge scholars and were asked to strictly follow these rules, I expect that voters have an even broader perception of pledges. To test this assumption, future research should analyse what types of sentences voters perceive as pledges. Experimental studies provide evidence that persuasive wordings and the specific statements slightly influence voters' perceptions of pledges (Lindgren 2018; Dupont et al. 2019). The analysis in Chapter 2 demonstrated that crowd workers generally have a broader understanding of pledges, even though these coders received the instructions developed by pledge scholars and were continuously monitored. Voters, who do not know about the scholarly conceptualisation, might have an even broader perception than the crowd workers. Survey experiments, which present coding instructions to one group while a control group does not receive any prior information, could enhance our understanding of the public definition of pledges. The fundamental differences between far-reaching scepticism and optimistic findings

in terms of pledge fulfilment require more research on citizens' understanding of election pledges.

6.3.3 The Development, Framing, and Diffusion of Pledges

Third, further research should analyse the framing and emphasis of election promises in the media. While I provide robust evidence for the assumption that newspapers focus more on broken and fulfilled promises, we do not know the policy areas of promises that receive more media attention. Only a small proportion of pledges seems to be covered in the media before elections. Topic models (Blei et al. 2003; Roberts et al. 2014) and supervised classifiers, trained on hand-coded party manifestos (Benoit et al. 2016; Merz et al. 2016), could identify the salient pledges prior to an election. Afterwards, researchers could estimate pledge fulfilment based on the promises that receive public attention.

Moreover, we require better knowledge of how pledges develop during an electoral cycle, and how promises diffuse across manifestos and in the media. Recent studies on party policy diffusion mostly focus on a specific policy, or measure diffusion using aggregated left-right positions (e.g., Böhmelt et al. 2016; Gilardi et al. 2018). Applying text similarity and 'text reuse' methods (Wilkerson et al. 2015; Linder et al. 2018) to national and subnational manifestos and newspaper articles from several countries will offer unique opportunities to study how pledges emerge and diffuse within a country, across countries, and during an electoral cycle. Advances in machine translation of political documents (Lucas et al. 2015; De Vries et al. 2018) hopefully enable the analysis of pledge diffusion across countries with different languages.

6.3.4 Crowd-Sourced Coding and Machine Learning

Finally, future projects should improve the combination of crowd-sourced coding and machine learning of political text. Chapter 3 demonstrates that crowd-sourcing

coding and machine classification can uncover previously neglected, but theoretically and practically important dimensions of political communication and party strategies. Crowd workers can already be employed to code political content from text, but for reasons of time and expense, coding a very large corpus of interest is infeasible. Therefore, future studies should analyse, in more detail, what types of machine learning methods perform best to correctly identify a crowd-coded dimension of interest. Researchers would also benefit from practical guidelines about the number of documents and coders required to apply the crowd-coded estimates to a large corpus.

Natural language processing, quantitative text analysis, and 'big data' approaches have found their way into social science research (Grimmer and Stewart 2013; Wilkerson and Casas 2017; Lazer and Radford 2017; Salganik 2017). I successfully applied quantitative text analysis and crowd-sourcing to a research area that is currently dominated by human coding. Nonetheless, the analyses on media reports and temporality of political statements should be extended by applying alternative classifications. Despite extensive validation and a theoretically justified keyword selection (Muddiman et al. 2018), the dictionary-based classification of media reports should be reassessed with deep learning methods, trained on a large hand-coded sample of promise-related statements. I hope that this dissertation encourages and inspires scholars to consider quantitative text analysis and natural language processing when trying to answer questions about representation and accountability.

6.4 Overall Implications for Democracy

Representative democracy requires a linkage and constant feedback between voters and parties. Yet, many citizens do not believe that politicians act as trustees or delegates, represent their voters' interests and implement the promised policies (Clarke et al. 2018). The rise of populist parties and party leaders is a logical

consequence of this development. This dissertation highlighted that researchers need to move beyond established approaches to analyse the way representative democracy works. Some of the basic assumptions of promissory representation (Mansbridge 2003) obviously do not hold in practice: many voters are not aware of fulfilled and broken promises and do not have an accurate understanding of promises for the upcoming cycle.

Focussing only on party manifestos to measure mandate fulfilment does not reflect reality. Alternative, and more visible sources of campaign communication should be considered. Otherwise, research might not grasp the promises that voters notice. In order to figure out why voters do not believe that parties fulfil their promises, we need to take a step back. Do voters notice or appreciate parties' attempts to claim credit? What policy or valence factors determine retrospective or prospective vote choice? What statements are considered pledges? And do parties learn from each other when developing campaign strategies and making promises? Answering these questions requires innovative methods and alternative data sources, but could equip researchers and practitioners with strategies to counteract legitimacy problems of democracy.

Appendix A

Reassessing an Established Concept Through Crowd-Sourced Text Coding (Supporting Information)

A.1 Additional Analyses and Robustness Checks

Table A.1 reports results from a multilevel logistic regression which tests whether the perceived difficulty of coding a statement or the coder's country of residence have an effect on the probability that a statement is coded as a pledge. The results are used to plot the predicted probabilities in Figures 2.7a and 2.7b.

Table A.1: Predicting the evaluation of a statement as a pledge, conditional on perceived coding difficulty and the coder's country of residence

	Model 1
Difficulty: Easy	0.55*
	(0.25)
Difficulty: Difficulty	0.06
	(0.28)
Difficulty: Very difficult	−0.07
	(0.46)
Country: Canada	−0.63
	(0.41)
Country: United Kingdom	−0.38
	(0.37)
Country: Ireland	−0.87
	(0.68)
Country: United States	−0.74*
	(0.37)
AIC	1813.98
BIC	1869.82
Log Likelihood	-897.99
N	3660
Num. grops: Sentences	139

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Note: The multilevel logistic model includes random intercepts for each statement. Standard errors in parentheses.

A.2 Coding Instructions for Crowd Workers

Note: The following instructions have been presented to every person interested in participating in the coding exercise.

Overview

This task involves reading sentences from political texts and judging whether these sentences can be classified as election pledges (promises). The sentences you will be asked about come from political party manifestos. A subset of the sentences contains election pledges. For the sentence highlighted in red, enter your best judgement about whether it fulfils the requirements to be defined as a pledge. If you are not entirely sure about the context of the highlighted sentence, read the surrounding sentences as well. Yet, your judgement should focus on the sentence in red font.

Below we define what is meant by a "pledge", and how to distinguish between different types of pledges.

1. What is an election pledge?

During election campaigns, political parties formulate election pledges. **An election pledge is a statement committing a party to one specific action or outcome that can be clearly determined to have occurred or not.**

First, we provide some examples of statements that should not be coded as a pledge.

No pledge

The following statements would be coded as "no pledge". The sentences are general descriptions of basic values and attitudes or statements of fact.

- "Our government should show respect for families." (This sentence is too vague, we cannot test in a meaningful way whether a government "showed respect".)
- "We are committed to environmental protection." (This statement is too generic. There is no concrete action promised to protect the environment – therefore this statement should not be coded as a pledge.)
- "Middle-class workers will continue to form the basis of our economy." (The sentence is a statement of 'fact', a description of the future. There is no concrete policy-action to be taken and promised.)

When you do code a statement as "no pledge", you will be asked (1) whether the sentiment of the statement is negative, natural, or positive, and (2) whether the statement refers to the past, the present situation, or the future.

Examples: Pledge

These statements should be coded as "pledge" because they contain an unambiguous, explicit and testable commitment for future actions.

- "We will set aside 1 percent of GNP to provide for future pension obligations."
- "We will establish a new National Development Finance Agency."
- "The party will work to achieve the situation where 80 percent of taxpayers pay only the standard tax rate."

Keep in mind that if one could equally strongly argue that an action was carried out or not, then you should not code the statement as a pledge. For instance, the sentence "We will do more to build pride in our institutions and to project our strength abroad." should not be coded as a pledge. It is impossible to test whether the actions of "building pride" and "project our strength abroad" have been carried out.

A pledge can also be part of a enumeration and does not have to be a complete sentence. For example, the sentence "To secure an additional 780 inward investment projects through a new fund." should be coded as a pledge. Even though a verb about that indicates future action ("will", "going to" etc) is missing, the statement contains and an unambiguous, explicit and testable commitment.

If you code a statement as "pledge", you will need to evaluate three additional features of the statement.

2. How precise is an election pledge?

Please determine the preciseness of the election promise on a scale from 1 (very vague) to 10 (very precise). *Very vague* pledges are imprecise, unspecific, and could be implemented in many ways. *Very precise* pledges promise an unambiguous course of action and can only be fulfilled in one way. The writers of the manifestos state very clearly what action will be taken.

Example: Very vague

- "We will conduct significant reforms in the tax credit system to effectively tackle current challenges."
- "We will improve the infrastructure to make regional locations more attractive."

These statements should be coded as "very vague". While the sentences expresses to change some aspects of the tax credit system and reform the infrastructure, there are many ways how a party could "conduct significant tax credit reforms" or "improve infrastructure". Keep in mind, however, that even very vague pledges need to outline a concrete policy action or outcome.

Example: Very precise

- "We are going to launch a Small Business Task Force focusing on farmers."
- "We complete the reduction of the standard rate of corporation tax to 12.5 percent by the end of the term."

The highlighted texts should be coded as "very precise" because there is only one way of fulfilling the promise – by launching the "Small Business Task Force focusing on farmers" or a "reduction of corporation tax to 12.5 percent".

3. What is the scope of an election pledge?

Finally, please determine the scope of an election pledge. The scope depends on how many people or groups of society will be affected by a pledge. Some election promises are relevant for many people and have more direct consequences than others. The scope ranges from 1 (very narrow) to 10 (very broad).

Example: Very narrow scope

- "The officials of health insurance agencies will be elected, not appointed."

The statement should be coded as "very narrow scope". The proposed policy (election of officials) only concerns people engaged in these agencies.

Example: Very broad scope

- "We will introduce free health insurance for all citizens."

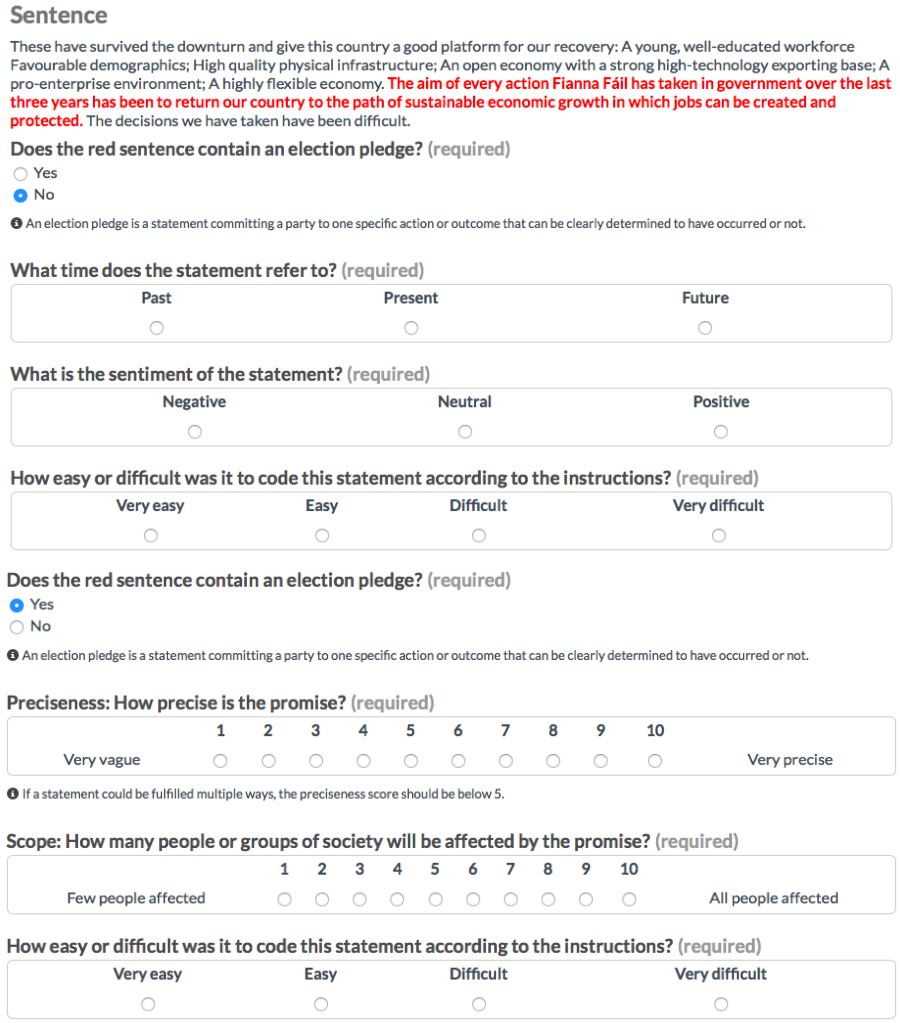
This statement should be coded as "very broad scope" because almost all citizens are likely to affected by the promised policy action ("the reform of the healthcare system").

4. How difficult was it to code the statement?

Finally, we ask you how easy or difficult it was to code the sentence according to the rules. Note that the answer to this question does not affect your performance rankings. Please answer this question honestly as it is very important to know which types of sentences are harder to code.

Last but not least, feel free to take part in the satisfaction survey and/or send me comments and feedback. Thank you very much!

Figure A.1: Flowchart of the CrowdFlower coding procedure



Note: The screenshot shows the workflow of the coding task. First, a respondent needs to decide whether the statement contains a pledge. If s/he does not code the statement as a pledge, the respondent will be asked about the tense and sentiment. If the statement is coded as a pledge, preciseness and scope need to be evaluated on a scale ranging from 1 to 10. For all statements, a respondent needs to evaluate the difficulty of coding the statement according to the instructions provided.

A.3 Aggregated Expert and Crowd Judgements

The table below lists all sentences with disagreement within the groups of experts *and* the crowd workers. These statements can be regarded as 'disputed' since at least one expert *and* one crowd worker deviate from the coding decision by the majority. The column 'Pledge mean' columns show the average pledge evaluation by the pledge scholars and the crowd workers. Recall that each statement was coded by at least 20 crowd workers and 9 pledge scholars. A value of 0 indicates that all crowd workers/experts coded the statement as 'no pledge', a value of 1 indicates that all crowd workers/experts coded a statement as 'pledge'. A value of 0.5 indicates the highest possible disagreement. The table below lists the text and average pledge evaluations for all 138 statements from the reliability exercise.

	Statement	Pledge mean (experts)	Pledge mean (crowd)
1	In a time of global economic uncertainty, Canada needs a strong and steady leader at the helm.	0.00	0.07
2	And we are defending our sovereignty over the vast Canadian Arctic.	0.00	0.00
3	We will make sure that Canada secures its strategic resources, that our energy development is clean and environmentally sustainable and that Canadians will be the main beneficiaries of the development and export of our resources.	0.00	0.91
4	Reaffirming the Ban on the Bulk Transfer of Water We will reaffirm the Canadian Government's position that NAFTA cannot require Canada to export bulk water to other NAFTA countries.	0.56	0.92
5	We will work with the provinces and territories to strengthen the current ban and ensure a clear legislative prohibition on bulk water removals or export from Canadian drainage basins.	0.78	1.00
6	Promoting the Development of Northern Pipelines We will help reduce regulatory and other barriers to increasing our pipeline network in the Far North in order to bring oil and gas to markets in Canada and throughout the world.	0.22	1.00
7	Prohibiting the Export of Raw Bitumen to Higher Polluting Jurisdictions We will prevent any company from exporting raw bitumen (unprocessed oil from the oil sands) outside of Canada for upgrading in order to take advantage of lower pollution or greenhouse gas emissions standards elsewhere.	1.00	1.00
8	Promoting Biofuels, Wind and Other Alternatives The Conservative Government is investing $1.5 billion over the next seven years in the production of biofuels, and requiring gasoline to contain 5 per cent renewable content by 2010 (2 per cent renewable content for diesel by 2012).	0.33	0.78
9	We are investing $1.48 billion over four years into incentives to produce more wind, solar, geothermal and tidal power - an investment that will produce over 14 million megawatt hours of clean electricity, or enough to power a million homes.	0.11	0.75
10	We will continue to make biofuels and renewable energy a key part of our environmental and energy strategy.	0.00	1.00

	Statement	Pledge mean (experts)	Pledge mean (crowd)
11	Generating 90% of Our Electricity From Non-Emitting Sources by 2020 As part of our Turning the Corner action plan, a re-elected Conservative Government will work to ensure that 90 per cent of Canadian electricity needs are provided by non-emitting sources such as hydro, nuclear, clean coal or wind power by 2020.	0.56	0.95
12	We will continue, as major priorities, to protect the sovereignty and promote the development of Canada's Arctic and North.	1.00	0.87
13	Today, Canada is strong, united and free - the envy of the world.	0.00	0.00
14	A re-elected government will: Introduce amendments to the Arctic Waters Pollution Prevention Act to extend our jurisdiction over polluting vessels to the edge of Canada's 200 mile exclusive economic zone.	0.00	1.00
15	Require mandatory notification of any foreign vessels entering Canadian territorial waters.	1.00	1.00
16	Assert Canada's rights over our Arctic waters, including the Northwest Passage, and work actively to complete the submission, by 2013, of scientific data on the extent of Canada's continental shelf under the United Nations Convention on the Law of the Sea.	0.89	1.00
17	Build a world-class High Arctic Research Station that will be on the cutting edge of Arctic issues, including environmental science and resource development.	1.00	1.00
18	The Conservatives and Stephen Harper believe that the current Senate must be either reformed or abolished.	0.22	0.08
19	An unelected Senate should not be able to block the will of the elected House in the 21st century.	0.00	0.03
20	As a minimum, a re-elected Conservative Government will reintroduce legislation to allow for nominees to the Senate to be selected by voters, to provide for Senators to serve fixed terms of not longer than eight years, and for the Senate to be covered by the same ethics rules as the House of Commons.	1.00	1.00
21	We will introduce legislation to move closer towards representation by population in the House of Commons for Ontario, British Columbia and Alberta, while protecting the seat counts of other provinces.	1.00	1.00
22	We will continue to reform appointments to federal agencies, boards, commissions and Crown corporations.	0.00	1.00
23	We will establish a task force to report within one year on unnecessary federally appointed positions that can be eliminated, with a target of reducing federal appointments by 10 per cent overall.	0.67	1.00
24	We are in the best financial position of any country in the G7.	0.00	0.00
25	We will work to ensure that appointees to federal agencies, boards, commissions and Crown corporations reflect the diversity of Canada in language, gender, region, age and ethnicity.	0.33	0.94
26	We will appoint members to the Public Appointments Commission.	1.00	0.79
27	The Commission will oversee the selection process for appointments to federal agencies, boards, commissions and Crown corporations, as well as develop guidelines, review and approve the selection processes for appointments and report publicly on the Government's compliance with the guidelines.	0.11	0.85
28	While the creation of the Commission was authorized by the Federal Accountability Act brought in by the current government, the previous Parliament blocked the appointment of several distinguished Canadians from sitting on the Commission.	0.00	0.08
29	A re-elected Conservative Government will ensure that the Public Appointments Commission gets up and running.	0.00	0.96
30	We will require all federal departments and agencies to produce detailed quarterly financial statements.	1.00	1.00
31	We will implement legislation, obstructed and delayed in the previous Parliament, which would prevent candidates for federal political office from taking out large private loans on non- commercial terms.	1.00	1.00
32	We will respect the jurisdiction of the provinces and territories in the Constitution Act, 1867, and will enshrine our principles of federalism in a new Charter of Open Federalism.	0.11	0.91

	Statement	Pledge mean (experts)	Pledge mean (crowd)
33	We will ensure that any new shared-cost program in an area of provincial or territorial responsibility has the consent of the majority of provinces to proceed, and that provinces should be given the right to opt out of the federal program with compensation, so long as the province offers a similar program with similar accountability structures.	1.00	1.00
34	We will respect the unique needs of French-language broadcasting services in Quebec by guaranteeing alternating French- and English-speaking Chairpersons of the CRTC and that one of the two Vice-Chairpersons will be from each official language group.	1.00	1.00
35	While other countries are struggling with uncertainty, Canada is on the cusp of becoming a bastion of economic strength on the international stage.	0.00	0.00
36	At least 25 per cent of the CRTC commissioners will be French-language speakers, appointed in consultation with Quebec and groups representing linguistic minorities.	0.89	0.57
37	Hearings related to Frenchlanguage or Quebec broadcasters will be heard by panels consisting of a majority of French-language or Quebec CRTC members.	0.67	0.37
38	We will increase funding for TV5, the international French-language television network, by $25 million over the next five years.	0.89	1.00
39	We will maintain financial support for arts and culture at or above existing levels, while continuing to improve the effectiveness of allocations wherever possible.	0.00	1.00
40	We will continue our record of strong support for the arts.	0.00	0.79
41	We will create a new, refundable tax credit on up to $500 of eligible fees for children under 16 who participate in eligible arts or cultural activities, such as music lessons, drama or art classes.	0.89	1.00
42	A re-elected Conservative Government will not reintroduce the Bill C-10 proposals to change film and video tax credit eligibility.	1.00	1.00
43	Although these proposals were approved unanimously by the House of Commons, we will take into account the serious concerns that have been expressed by film creators and investors.	0.00	0.59
44	We will take steps to help celebrate Canadian history and identity and develop a stronger sense of national citizenship.	0.11	0.95
45	Despite our gains over the past two and a half years, Canada is at risk of going back.	0.00	0.00
46	This year, Canada celebrates the 400th anniversary of the establishment of Quebec City, which marked the foundation of the Canadian state.	0.00	0.00
47	As Canada approaches other important historical milestones, such as the 400th anniversary of our first English settlement in Cupids, Newfoundland and Labrador in 2010, the 200th anniversary of the War of 1812 and the 150th anniversary of Confederation in 2017, a Conservative Government will act to ensure that these occasions achieve the national recognition they deserve.	0.00	0.79
48	We will continue support for sport and amateur fitness in Canada, at both the elite and recreational levels.	0.11	0.85
49	We will also enhance the Children's Fitness Tax Credit by making it refundable.	0.89	1.00
50	A Harper Conservative Government will continue to guide the country on a steady course through global instability.	0.00	0.46
51	We will continue to stand up for Canada's national interests as a credible competitor and attract better and higher-paying jobs.	0.00	0.75
52	We will deliver on a modest and realistic plan that includes making it more affordable for a young couple to buy their first home, for parents to save for their children's education and for seniors to retire to a comfortable life.	0.00	0.80
53	We will continue to make our communities safer by ending house arrest for serious and violent crimes, including home invasion and drug trafficking.	0.89	1.00
54	We will help youth at risk of gang activity and introduce a new law on young offenders that will impose stiffer sentences for crimes like murder and manslaughter.	0.78	1.00

	Statement	Pledge mean (experts)	Pledge mean (crowd)
55	For the past two and a half years, Prime Minister Stephen Harper has provided that leadership by managing the economy with prudence and discipline and by providing Canadian families and businesses with practical help that's making a real difference.	0.00	0.00
56	We will also protect consumers from illegal and unfair retail practices and protect consumers and children from unsafe imported goods, including toys and food.	0.00	0.96
57	Under Stephen Harper's leadership, Canada will emerge stronger and more united than ever before.	0.00	0.34
58	This election, Canadians face a clear choice.	0.00	0.00
59	It's a choice between the Harper Conservatives' credible and affordable plan, and risky tax-and-spend experiments that will drive up the cost of everything from groceries to gas and throw Canada back into a deficit.	0.00	0.04
60	It's a choice between strength and weakness.	0.00	0.00
61	And it's a choice between moving forward and going back.	0.00	0.04
62	As high energy costs and turbulence in international financial markets have taken their toll on the world economy, Canadians have become increasingly concerned about the rising cost of living and threats to our quality of life.	0.00	0.00
63	Stephen Harper and the Conservative Government have helped Canadians cope with these challenging times.	0.00	0.00
64	Reducing the GST by two per cent has helped boost economic growth during the current slowdown.	0.00	0.00
65	This is part of a plan that has helped the Canadian economy.	0.00	0.07
66	Canada has witnessed the net creation of more than 800,000 new jobs, while the Harper Government has lowered taxes for the typical family by more than $3,000 a year and relieved our children of nearly $40 billion in national debt.	0.00	0.00
67	Savings and benefits are more than $3,000 per year for a typical, two-child Canadian family earning $80,000.	0.00	0.00
68	A re-elected Conservative Government will continue to provide practical help to Canadian families to assist them with higher costs of living, and protect them from unfair retail practices so that families can focus on the things in life that matter most, like buying their first home and saving for their children's education.	0.00	0.85
69	We will reduce the federal excise tax on diesel and aviation fuel by half - from four cents per litre to two cents per litre - reducing the price of transportation by truck, train, plane and ship, and helping to bring downward pressure on consumer prices.	0.89	1.00
70	We will implement a strong consumer protection plan by modernizing Canada's outdated competition laws.	0.00	0.94
71	New competition laws will: Make it easier to investigate and prosecute bid-rigging and hard-core cartel behaviour such as price fixing.	0.33	0.70
72	Raise maximum penalties for bid-rigging and cartels to a $25-million fine and 14 years in prison.	0.89	0.93
73	Introduce fines of up to $10 million - $15 million for repeat offenders - for companies that abuse their dominant market position.	0.89	0.89
74	Provide for restitution for consumers who fall victim to deceptive marketing practices.	0.56	0.78
75	We will prevent telecommunications companies from charging fees to customers for receiving unsolicited commercial text messages.	0.78	1.00
76	We will amend the Telecommunications Act to strengthen the power of the Commissioner of Complaints for Telecommunications Services, including the creation of a code of conduct for wireless services.	0.78	1.00
77	People who work hard, pay their taxes and play by the rules are getting ahead.	0.00	0.00
78	We will also create a compliance and deterrent power that allows the Canadian Radio- television and Telecommunications Commission (CRTC) to block these and similar unfair charges in the future.	0.56	1.00
79	We will introduce legislation to prohibit the use of spam (unsolicited commercial email) to collect personal information under false pretences and to engage in criminal conduct.	0.89	1.00

	Statement	Pledge mean (experts)	Pledge mean (crowd)
80	The new law will reduce dangerous, destructive and deceptive email and web site practices, and will establish new fines for those who break the law.	0.44	0.75
81	We will extend our new truth in labelling guidelines for food products to consumer products, ensuring that Canadians know more about the products they purchase.	0.22	1.00
82	As was done with food products, the use of the "Product of Canada" label will be restricted to consumer products where both the contents and processing are Canadian.	1.00	0.83
83	Qualified "Made in Canada" labels will be used for consumer products that are processed in Canada, but contain imported content, such as "Made in Canada from imported contents."	0.89	0.83
84	We will follow through with consumer product safety legislation providing for more inspection and testing of food, toys and consumer products to ensure that dangerous products are swiftly recalled.	1.00	1.00
85	The legislation will also establish stiff offences and penalties for violations.	0.56	0.90
86	We will ensure that any regulation of natural health products balances the protection of Canadians' health and safety with the freedom to choose alternative products.	0.11	0.84
87	Establishing Tough Laws on Tampering with Gas Pumps We will amend the Weights and Measures Act and the Electricity and Gas Inspection Act to ensure the accuracy of gas pumps and home-heating meters.	0.89	0.95
88	The Harper Government has taken bold action to make Canada a safer place for families too.	0.00	0.00
89	We will increase fines for companies that overcharge consumers, with additional fines as high as $50,000 for repeat offenders.	0.78	1.00
90	We will allow families where one spouse is not working full-time in order to care for one or more family members with disabilities - whether children or adults - to split their income between spouses for tax purposes.	1.00	0.97
91	We will make it easier for families to plan for the future of their children with disabilities by improving the Disability Savings Plan.	0.00	1.00
92	We will allow the proceeds of a deceased individual's Registered Retirement Savings Plan (RRSP) or Registered Retirement Income Fund (RRIF) to be rolled over on a tax deferred basis to the Disability Savings Plan of a financially dependent infirm child or grandchild.	0.89	1.00
93	This builds on the Disability Savings Plan established by this government in 2007, and will make it easier for a person with disabilities to access money that has been transferred from the unused retirement savings of a deceased family member.	0.00	0.68
94	We will give first- time homebuyers a tax credit for up to $5,000 of eligible closing costs on the purchase of a new home.	1.00	1.00
95	A re-elected Conservative Government will expand the current Registered Education Savings Plan (RESP) to allow charities and not-for-profit organizations to establish RESPs, in partnership with parents, for children from low-income families that otherwise would not have the ability to save.	1.00	0.92
96	We will fully index the $100 per month Universal Child Care Benefit to inflation.	1.00	1.00
97	This will ensure that the value of the benefit does not erode over time.	0.00	0.38
98	To help sole-support, single-income parents with the costs of caring for children under six years old, a re-elected Conservative Government will make the $100 per month per child Universal Child Care Benefit given to these parents tax free.	1.00	1.00
99	Our tough-on-crime approach has resulted in tougher penalties for serious drug crimes and drug-impaired driving, and better protections from gangs, dangerous offenders and sexual predators.	0.00	0.00
100	This will recognize the uniquely high child care burdens borne by single parents.	0.00	0.16
101	We will increase the Senior Age Credit Amount by an additional $1,000, on top of already planned increases.	1.00	1.00
102	Once this increase is fully implemented, it will result in savings of more than $400 per year for eligible seniors receiving the full value of the credit.	0.33	0.61

	Statement	Pledge mean (experts)	Pledge mean (crowd)
103	We will end the unfair treatment of Canadians receiving U.S. Social Security payments.	1.00	1.00
104	Starting in 1997, the Liberal government began to tax 85 per cent of U.S. Social Security payments received by Canadian seniors.	0.00	0.04
105	A re-elected Conservative Government will restore the 50 per cent inclusion rate that applied before the Liberals took office.	0.00	0.97
106	The restored, lower rate will apply to Canadians who were receiving U.S. Social Security payments prior to January 1, 1996.	0.11	0.65
107	National unity is stronger than it has been in forty years.	0.00	0.00
108	Stephen Harper and the Conservatives are committed to ensuring that all Canadians have an adequate quality of housing.	0.00	0.51
109	We have extended the Homelessness Partnering Strategy for the next two years and will maintain funding to deal with homelessness at least at the current levels.	0.78	0.79
110	We will work with provinces, municipalities and charitable organizations to look at ways of improving the effectiveness of federal dollars spent on homelessness initiatives.	0.00	1.00
111	We have also renewed funding for the Residential Rehabilitation Assistance Program to assist low-income Canadians in renovating their homes and for the Affordable Housing Initiative, which supports the creation of new rental housing units.	0.11	0.18
112	We will ensure that Aboriginals have the opportunity to fully participate in Canada's economy and society.	0.00	0.82
113	Improving Aboriginal education is crucial to giving young members of the Aboriginal community the opportunity to succeed.	0.67	0.43
114	We will work to complete tripartite educational agreements, modelled after the successful agreements in British Columbia and New Brunswick, with provinces and First Nations organizations across the country.	0.00	1.00
115	A re-elected Conservative Government will also commit to pursuing bilateral agreements with provinces to address the wrongs of the residential schools era for Aboriginals attending similar schools not covered by the Indian Residential Schools Settlement Agreement.	0.67	0.89
116	Since coming to office, Stephen Harper and the Conservatives have taken steps to help Canada compete in the world economy and to make sure that we are creating and preparing Canadians for the next generation of work.	0.00	0.05
117	And Canada is, once again, taking our proper place on the world stage.	0.00	0.00
118	We reduced taxes on business, invested in science and technology and provided incentives for young Canadians to enter the skilled trades.	0.00	0.07
119	We took measures to encourage more participation in the workforce, such as the $1,000 Canada Employment Credit, the Working Income Tax Benefit for lower- income working families and efforts to attract more skilled workers through immigration and to promote recognition of international credentials.	0.00	0.05
120	We will do more to invest in jobs for the future and for a flexible labour market that responds to the needs of Canadians.	0.00	0.92
121	We will give self- employed Canadians the opportunity to access maternity and parental benefits.	0.89	1.00
122	We will continue to reduce taxes for small and medium-sized businesses.	0.11	1.00
123	In our first term, we reduced the small business tax rate to 11 per cent, raised the eligibility threshold for the small business income tax rate to $400,000 and raised the lifetime capital gains exemption for small business owners to $750,000.	0.00	0.10
124	We will build on this record in our second term by: Raising the small business eligibility threshold to $500,000.	1.00	1.00
125	We will build on this record in our second term by: Indexing the lifetime capital gains exemption to inflation.	1.00	1.00
126	We will achieve its target of a 20 per cent reduction this year in reporting requirements for small businesses and will continue to reduce the burden of unnecessary red tape and paperwork on Canada's small business sector.	0.67	0.96

	Statement	Pledge mean (experts)	Pledge mean (crowd)
127	A re-elected Conservative Government will continue to work with the private sector and small business organizations to identify inefficiencies and make permanent a formal process of measuring, reporting and reducing the burden on businesses.	0.56	0.94
128	We are promoting our national interests and spreading our Canadian values of freedom, democracy, human rights and the rule of law.	0.00	0.04
129	We will create a new $75- million venture capital fund, to be administered by the Business Development Bank of Canada, which will allow late-stage technology companies to move from research and development to commercialization stages of business.	1.00	1.00
130	We will reintroduce federal copyright legislation that strikes the appropriate balance among the rights of musicians, artists, programmers and other creators and brings Canada's intellectual property protection in line with that of other industrialized countries, but also protects consumers who want to access copyright works for their personal use.	0.78	1.00
131	We will also introduce tougher laws on counterfeiting and piracy and give our customs and law enforcement services the resources to enforce them.	0.78	1.00
132	This will protect consumers from phoney and sometimes dangerous products that are passed off as reliable brand-name goods.	0.00	0.45
133	After two and a half years with Stephen Harper and the Conservatives, our country is stronger and more united than it has been in more than 40 years.	0.00	0.00
134	Support for federalism in Quebec is at a modern-day high.	0.00	0.05
135	We have cleaned up lobbying and campaign financing and have begun to reform our democratic institutions.	0.00	0.15
136	We have started to rebuild our national defence and to play a bigger role on the world stage.	0.00	0.00
137	Our abundant natural resources, especially in our vast, untapped Arctic, have become key strategic assets as the world focuses more on energy and the environment.	0.00	0.00
138	A re-elected Conservative Government will do more to build pride in Canada's institutions and to project our strength abroad.	0.00	0.47

Appendix B

Prospective and Retrospective Rhetoric: A New Dimension of Party Competition and Campaign Strategies (Supporting Information)

B.1 Classifying Retrospective and Prospective Manifesto Sentences

The English classifier was trained on a crowd-coded corpus consisting of the 2011 manifesto of the Irish party Fianna Fáil. The German classifier was trained on the sentence-level annotation of all 2013 Austrian manifestos which was part of the Austrian National Election Study (Müller et al. 2017).[1] The sentences in the English and German training set contain a label that indicates whether a statement was classified as 'retrospective' or 'prospective'.

[1]The AUTNES coding contained the categories 'no pledge', 'subjective (not testable) pledge', and 'pledge'. I combined sentences from the latter two categories to the 'prospective' category; all other sentences were coded as 'retrospective'. Each sentence from the Austrian manifestos was coded by two trained researchers (Dolezal et al. 2016).

To train the English Naïve Bayes classifier, I relied on crowd-sourced annotation of the entire 2011 Fianna Fáil manifesto. The test sets for cross-validation were also crowd coded with the same procedure. The coding was preceded by five coding jobs and continuous feedback to test and improve the coding instructions. In the following, I describe the coding process, quality control mechanisms as well as the number of codings and aggregation.

Coding Process and Instructions

Workers are recruited and data are collected on the online platform CrowdFlower (renamed to Figure Eight in 2018).[2] Potential contributors first read detailed coding instructions with a pledge definition and several examples (Appendix B.2). Workers are required to code strictly according to the instructions, not based on their personal definition of an election pledge. Afterwards, workers need to answer four out of five questions correctly. Having passed this quiz, respondents code randomly selected sentences from the text corpus in groups of five questions, and get paid $0.03 per coded statement which was 50 per cent higher than the default payment.

The coding process works as follows: Contributors must decide whether the statement should be coded as a pledge according to the definition provided in the instructions (Section B.2). Election pledges are "statements committing a party to an action or outcome that is testable", in the sense that researchers could gather evidence to ascertain whether or not the action or outcome was accomplished (Thomson et al. 2017: 532). If a statement is not classified as a pledge, the coder must decide whether the statement relates to the future (but does not consist of an election pledge) or whether the sentences is about the past or present. If a sentence contains more than one temporal dimension, the respondent should mark the statement as ambiguous and select the second tense category. Based on the recommendations by Benoit et al.

[2]https://figure-eight.com. Figure Eight has more than 10,000 registered workers from all over the world.

(2016), each statement is coded by at least five 'trusted' crowd workers who had an accuracy of at least 80 per cent correctly answered test questions (initial quiz not included). For the training set, I only selected sentences that were coded in the same way – in terms of the temporal focus – by at least four out of five crowd workers (which corresponds to 83 per cent of all coded sentences). The original coding consists of four categories: past, present, subjective pledge, objective (testable) pledge. The first two groups capture retrospective statements, while the last two groups summarise prospective statements.

Quality Control

Crowd sourced text analysis requires test questions (also called gold tasks) as a control system to remove 'spammers' (Benoit et al. 2016). For these test questions, the 'answer key' (how the statement should be coded correctly) is specified in advance. 20 per cent of all sentences are test questions occurring at a random position in each block of five statements to be evaluated. As test questions I selected a sample of statements from a reliability test of the 2002 Fianna Fáil manifesto and the 2008 manifesto of the Conservative Party of Canada that were coded identically by the nine expert coders from the Comparative Party Pledges Group (Thomson et al. 2017). For each of the 60 test questions I have aggregated and fully anonymised the scholarly coding. Workers need to answer 80 per cent of the test/screener questions correctly throughout the job – a benchmark based on existing studies (Benoit et al. 2016). As the workers' origin I chose English speaking countries (United Kingdom, United States, Republic of Ireland, Canada, Australian, New Zealand). Participants needed to have completed at least 100 test questions in earlier jobs and have had an overall of 80 per cent correctly answered test questions throughout their prior coding jobs. Section B.2 shows the coding instructions and answer options for an English coding task.

B.2 Coding Instructions for Crowd Workers

Note: The following instructions have been presented to every person interested in participating in the coding exercise.

Overview

This task involves reading sentences from political texts and judging whether the statement relates to the past, present situation, the future, or whether the statement contains an election promise (pledge). The sentences you will be asked about come from political party manifestos.

For the sentence highlighted in red, enter your best judgement about the temporal coverage. If you are not entirely sure about the context of the highlighted sentence, read the surrounding sentences as well. Yet, your judgement should focus on the sentence in red font. Below we define the categories and provide examples.

1. What is the temporal direction of a statement?

a) **Statements about the past** describe achievements, criticisms or facts that have happened in the past.

Examples:

- "The government has neglected to invest substantial resources to ensure that the country has major international connectivity capacity."
- "We have reduced waiting times for the collection of IDs and turnaround time for social grant applications."

b) **Statements about the present** situation describe, criticise or praise the current situation.

Examples:

- "In our democratic country, women's voices are heard and women's issues are seriously addressed."
- "We are in living in a time of high unemployment and poverty."

c) **An election promise** is a statement about the future that commits a party to one specific action or outcome. This outcome or action can be clearly determined to have occurred or not.

Examples:

- "We will set aside 1 percent of GNP to provide for future pension obligations."

- "We will establish a new National Development Finance Agency."
- "The party will work to achieve the situation where 80 percent of taxpayers pay only the standard tax rate."

If one could equally strongly argue that a clear action or outcome is promised, then you should not code the statement as a pledge.

d) **Statements about the future** describe actions or situations that might or will happen. In contrast to election promises, statements about the future do not outline a clear policy goal or outcome that a party commits itself to.

- "In the future we will stay committed to environmental protection." (This statement is about the future, but does not promise a concrete action or outcome.)
- "Middle-class workers will continue to form the basis of our economy." (The sentence is a description of the future. There is no concrete policy-action to be taken and promised.)

2. Ambiguous statements

Sometimes, a statement could belong to two categories (for example if it describes both the past and the present). If this is the case, please indicate that the statement could be classified into a second category and specify the temporal coverage.

B.3 Alternative Estimations of Sentiment and Face Validity

Figures B.1–B.3 plot the correlations between the aggregated sentiment scores on the level of manifesto-classes, using two different dictionaries (LSD and LIWC) and two different approaches for aggregating sentiment across sentences (Soroka 2012; Proksch et al. 2019).

Figure B.1: Correlations between English sentiment dictionaries

(a) Calculation based on Proksch et al. (2019) **(b)** Calculation based on Soroka (2012)

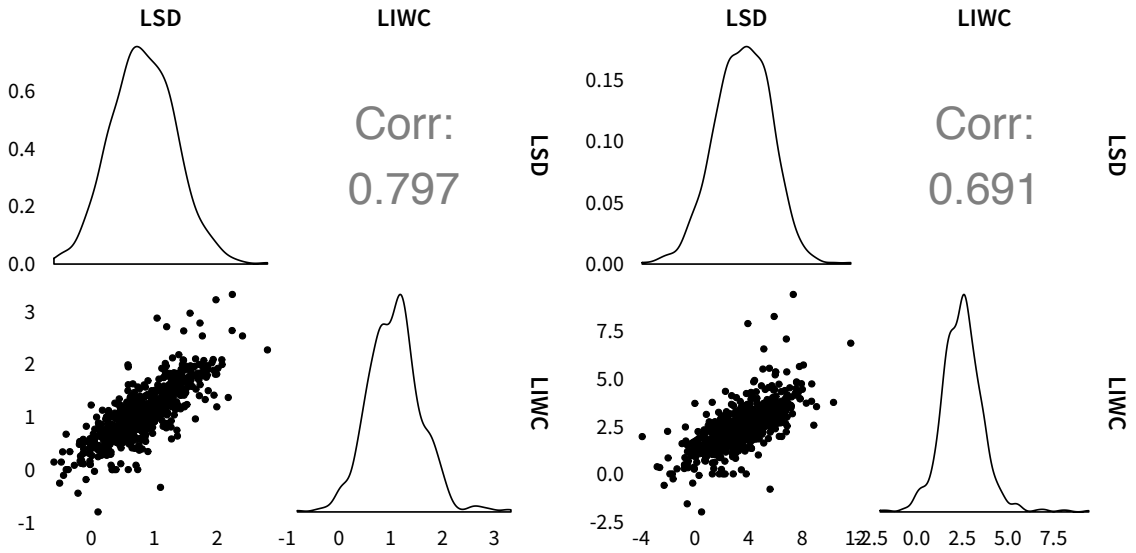


Figure B.2: Correlations between German sentiment dictionaries

(a) Calculation based on Proksch et al. (2019) **(b)** Calculation based on Soroka (2012)

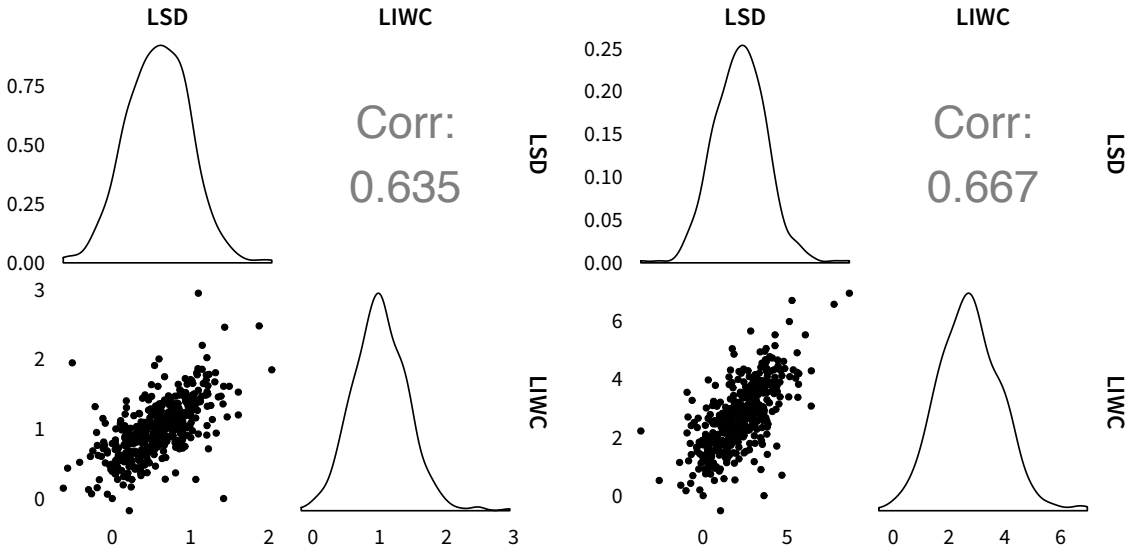
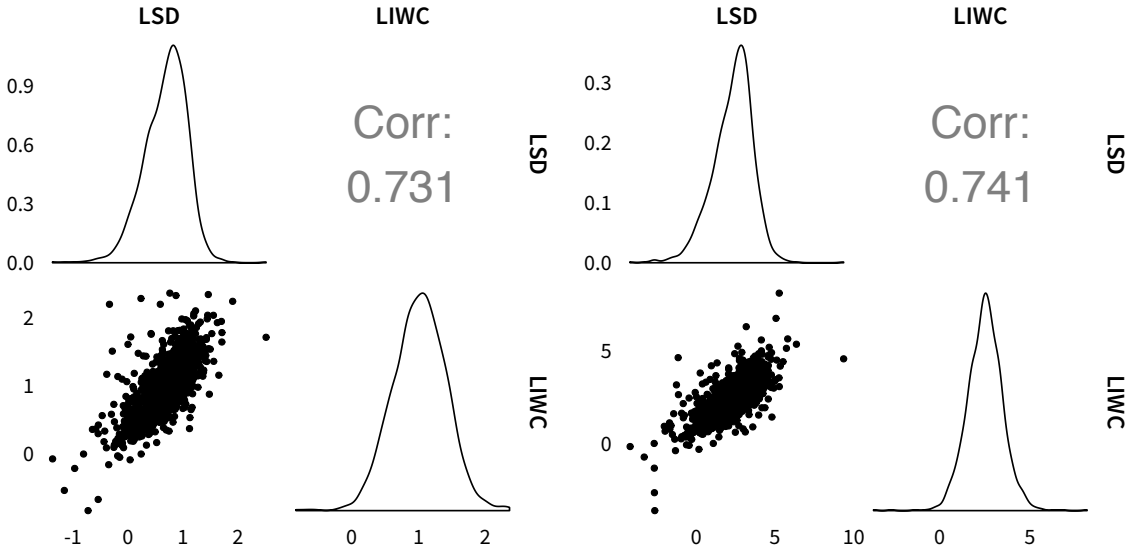


Figure B.3: Correlations between German sentiment dictionaries for German regional manifestos

(a) Calculation based on Proksch et al. (2019) **(b)** Calculation based on Soroka (2012)



Tables B.1–B.4 list examples of retrospective and prospective sentences from English and German manifestos from national general elections which are among the 50 most positive and negative sentences in each language and class (according to the Lexicoder Sentiment Dictionary and Proksch et al.'s (2019) aggregation).

Table B.1: Examples of English positive sentences

Sentence	Class	Party
We will protect the media from over-regulation; support an open and accountable media; and protect the freedom of speech to strengthen our vibrant democracy.	Prospective	Liberal Party of Australia (2013)
This confidence is essential for encouraging new investment, maintaining and creating jobs, and generating the national wealth necessary to assure Canadians a stable and secure future.	Prospective	Liberal Party of Canada (1997)
Enforcement of the law is the responsibility of a civil police force in which efficiency loyalty and respect for the rights of the individual will be encouraged and supported.	Prospective	NZ National Party (1966)
John G. Diefenbaker was a tireless advocate for freedom and human rights, and an early and powerful champion for Canada's North and Northern peoples.	Retrospective	Conservative Party of Canada (2015)
A strong education system supports a sustainable and adaptable society capable of thriving in the face of multiple changes.	Retrospective	Green Party (2016)
Expert opinion confirms what common sense tells us: children well taught and well-cared-for in their early years have a better opportunity to lead successful and rewarding lives.	Retrospective	Liberal Democrats (2005)

Table B.2: Examples of English negative sentences

Sentence	Class	Party
Putting the victim in fear of injury will be a serious aggravating factor in any offence against the person.	Prospective	Labour Party (2007)
Non-enforcement, insufficient penalties and deficiencies in law all contribute to high-levels of work-related illness, injuries and fatalities.	Prospective	Sinn Féin (2007)
We are tackling the problem of those who fraudulently pose as refugees and who seek to exploit Britain's long tradition of giving refuse to the victims of persecution.	Prospective	Conservative Party (1987)
We know the fear and intimidation suffered by the victims, who are often themselves vulnerable people.	Retrospective	Labour Party (2007)
Some aspects of the steadily growing crime rate are alarming, particularly senseless crimes of wanton destruction and hooliganism.	Retrospective	NZ Labour Party (1966)
They put an end to federal court early-release orders for prison overcrowding and made it much harder for prisoners to file frivolous lawsuits about prison conditions.	Retrospective	Republican Party (1996)

Table B.3: Examples of German positive sentences

Sentence	Class	Party
Um mehr Klimaschutz, mehr Gerechtigkeit, mehr Freiheit, mehr Menschenrechte und mehr Frieden zu erreichen, müssen wir die Idee Europas erneuern und die EU stärken.	Prospective	Alliance‘90/Greens (2009)
Wir fördern wichtige Programme wie „Soziale Stadt” oder „Demokratie leben”, die Stadtentwicklung, Sicherheit, Stärkung unserer Demokratie, Engagement und Hilfe für Jung und Alt verbinden.	Prospective	Social Democratic Party of Germany (2017)
Wir wollen Kinder mit Lernschwierigkeiten besser fördern, behinderte Kinder soweit wie möglich mit allen anderen zusammen erziehen, fördern statt auslesen bleibt Grundsatz unserer Bildungspolitik.	Prospective	Social Democratic Party of Germany (1980)
Eine zeitgemäße Schule bietet abwechselnd Unterricht und Freizeit, mit motivierten und bestens ausgebildeten LehrerInnen und FreizeitpädagogInnen, die Kinder umfassend fördern und unterstützen, wo sie es brauchen.	Retrospective	Austrian Social Democratic Party (2013)
Werte wie Ehrfurcht vor Gott, Wahrhaftigkeit, Ehe und Familie, Verantwortung, Menschenwürde und Gerechtigkeit gehören zum Abc von Gesellschaft und Politik.	Retrospective	Federal Democratic Union (2007)
In der Familie lernen die Menschen Tugenden und Verhaltensweisen, die unserer Gesellschaft ein menschliches Gesicht geben: Liebe und Vertrauen, Toleranz und Rücksichtnahme, Opferbereitschaft und Mitverantwortung.	Retrospective	Christian Democratic Union/Christian Social Union (1983)

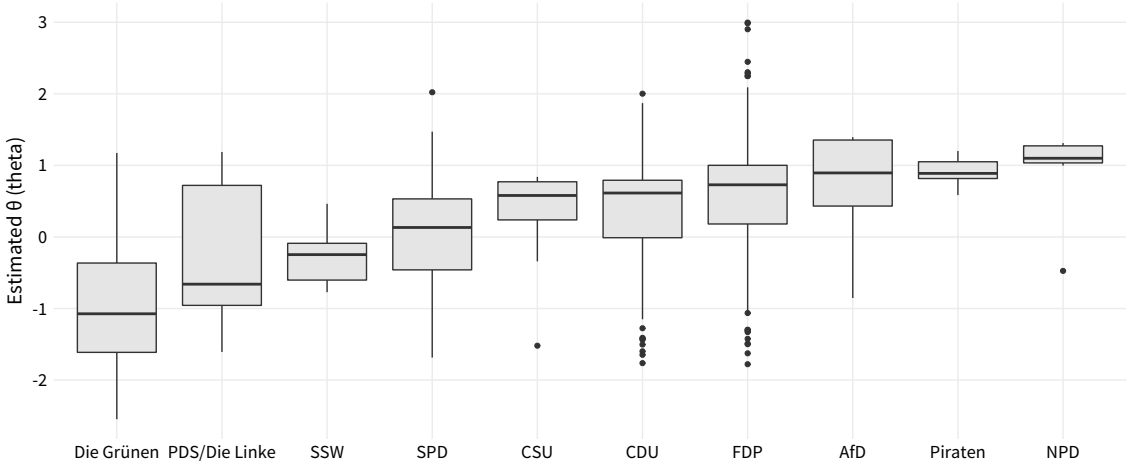
Table B.4: Examples of German negative sentences

Sentence	Class	Party
Um einen Wildwuchs an Forschungseinrichtungen zu vermeiden, hat der Staat – möglichst ohne direkte Eingriffe in Einklang mit den Betroffenen für ein entsprechendes Forschungskonzept zu sorgen.	Prospective	Austrian Freedom Party (1999)
Verstöße gegen Kennzeichnungsvorschriften oder gegen Richtlinien für Öko-Produkte werden mit Strafen belegt.	Prospective	Social Democratic Party of Germany (2002)
Ich werde Maßnahmen gegen den Leerlauf in den Spitälern und gegen die Erosion der Spitalskosten ergreifen.	Prospective	Austrian People's Party (1983)
Wer Angst haben muss um seine Existenz, vor Armut im Alter oder schlechter Versorgung bei Krankheit, der lebt in Unsicherheit.	Retrospective	Social Democratic Party of Germany (2005)
Wir sehen in der Praxis der Grenzschutzagentur FRONTEX und der EU-Mitgliedsstaaten einen Verstoß gegen das Verbot der Zurückweisung und eine schwere Menschenrechtsverletzung.	Retrospective	Alliance‘90/Greens (2013)
Große Teile der Weltbevölkerung leiden unter Armut und Hunger, Menschen sterben an behandelbaren Krankheiten, an mangelndem Zugang zu sauberem Trinkwasser.	Retrospective	The Left (2013)

B.4 Manifestos from German Subnational Elections

Figure B.4 plots the distributions of ideological positions of subnational party manifestos on a unidimensional scale, estimated through Wordfish scaling (Slapin and Proksch 2008; Lowe and Benoit 2013).

Figure B.4: The distribution of ideological positions of German parties



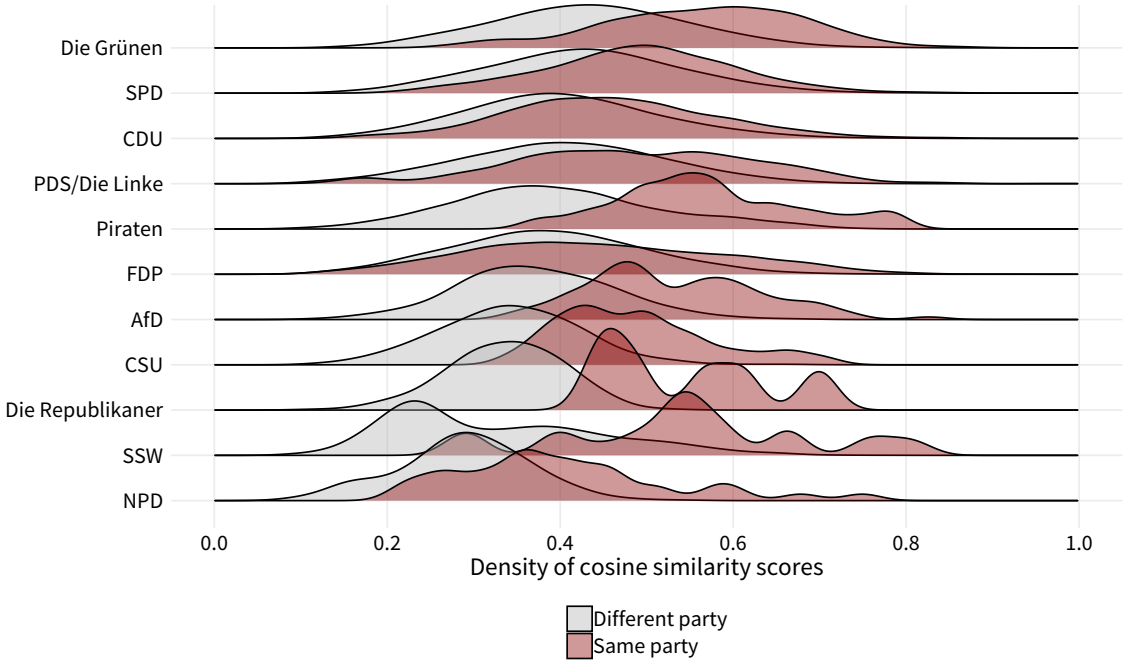
Note: The boxplots consider 631 party manifestos, scaled through Wordfish (Slapin and Proksch 2008). Parties with fewer than six manifestos are omitted from the plot.

Figure B.5 plots the density of cosine similarities between all 644 manifestos after removing 'stopwords' and punctuation. A value of 1 means that two manifestos share the exact same frequency of all terms. A value of 0 indicates no similarity between manifestos. The visualisation of all 402,590 similarity comparisons reveals an average cosine similarity of 0.4 (SD: 0.11) between manifestos from different parties, and a similarity of 0.48 (SD: 0.13) if two manifestos were written by the same party.[3]

While it is reasonable that manifestos from the same party have a higher similarity than from opponents, even the similarity values across the same party are rather low. Manifestos are usually customised for each election. I also ran linear regressions with the cosine similarity as the dependent variable (Table B.5). Model 1 only includes

[3]As a point of comparison, the 21 inaugural speeches of German chancellors at the beginning of each legislative cycle between 1949 and 2017 have an average cosine similarity of 0.57 (SD: 0.13).

Figure B.5: Distributions of cosine similarities, faceted by party



Note: The density curves show the distributions of cosine similarities both for manifestos drafted by the same party and for manifestos drafted by different parties. The plot considers only parties that contested at least three elections. The density curves are based on 402,590 similarity pairs.

a dummy indicating whether two manifestos were written by the same or different parties. This model confirms the descriptive evidence revealing that on average manifestos from the same party have a higher similarity of 0.08. However, when including dummies that check whether two manifestos were drafted by parties in the same Bundesland, the model fit improves considerably (the adjusted R^2 increases from 0.13 to 0.2). The coefficient for the same Land amounts to 0.13, the coefficient for the same election is 0.07, indicating that election- and Land specific topics are clearly considered by parties when drafting manifestos. Only 10 of the 644 manifestos have cosine similarities above 0.95, indicating that most parts of a manifestos have been copied from an earlier document (see Table B.6). All of these manifestos come from the same party in the same Land. Excluding manifestos that contain high proportions of reused text does not change the coefficients from the main model

(Table B.7). The analysis underscores that regional election manifestos are almost always carefully drafted for a specific election and allow us to test the hypotheses regarding parties' campaign communication conditional on incumbency status and office aspiration.

Table B.5: Predicting cosine similarities between party manifestos in German regional elections

	Model 1	Model 2	Model 3
Intercept	0.38***	0.37***	0.38***
	(0.00)	(0.00)	(0.00)
Same party	0.08***	0.08***	0.08***
	(0.00)	(0.00)	(0.00)
Same Bundesland		0.13***	
		(0.00)	
Same election			0.07***
			(0.00)
R^2	0.13	0.20	0.14
Adj. R^2	0.13	0.20	0.14
Num. obs.	207046	207046	207046
RMSE	0.12	0.11	0.12

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.
Note: The dependent variable is the cosine similarity between a pair of party manifestos. Party fixed-effects omitted from table. Standard errors in parentheses.

Table B.6: Pairs of subnational election manifestos with cosine similarities exceeding 0.95

First election	Second election	Party	Bundesland	Similarity
2011	2016	PDS/Die Linke	Baden-Württemberg	1.00
1986	1987	FDP	Hamburg	0.99
1982	1983	CDU	Hessen	0.99
2008	2009	FDP	Hessen	0.99
2008	2009	SPD	Hessen	0.99
1987	1988	SPD	Schleswig-Holstein	0.99
1982	1983	SPD	Hessen	0.98
1982	1983	Die Grünen	Hessen	0.97
1986	1987	CDU	Hamburg	0.96
2008	2009	PDS/Die Linke	Hessen	0.96

Table B.7: Predicting sentiment in party manifestos from German Länder elections (excluding heavily reused manifestos)

	Model 1 (main model)	Model 2 (exclude reused)
Opposition	-0.17^{***}	-0.17^{***}
	(0.03)	(0.03)
Class: Retrospective	-0.20^{***}	-0.20^{***}
	(0.02)	(0.02)
Seat share	0.34^{**}	0.33^{**}
	(0.12)	(0.12)
Left-right (Wordfish)	0.85^{*}	0.71
	(0.42)	(0.42)
Left-right^2 (Wordfish)	-1.66^{***}	-1.62^{***}
	(0.36)	(0.36)
Year	0.01^{***}	0.01^{***}
	(0.00)	(0.00)
Opposition × Class: Retrospective	-0.09^{***}	-0.09^{***}
	(0.02)	(0.02)
AIC	74.72	73.45
BIC	141.81	140.34
Log Likelihood	-24.36	-23.73
N	1288	1268
Num. groups: Manifesto	644	634
Num. groups: Election	141	141
Num. groups: Bundesland	16	16
Num. groups: Party	15	15

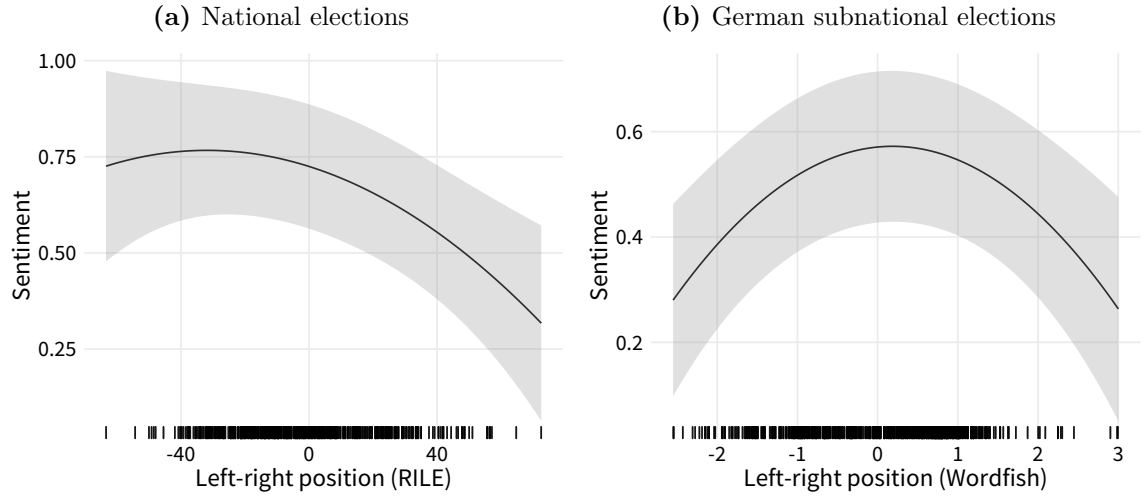
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Note: Multilevel linear regressions with the sentiment as the dependent variable. Model 1 is the same as Model 3 in Table 3.3. Model 2 excludes 10 manifestos that have very high cosine similarity values (>0.95) compared to a previously published manifesto. Standard errors in parentheses.

B.5 Additional Regression Models and Plots

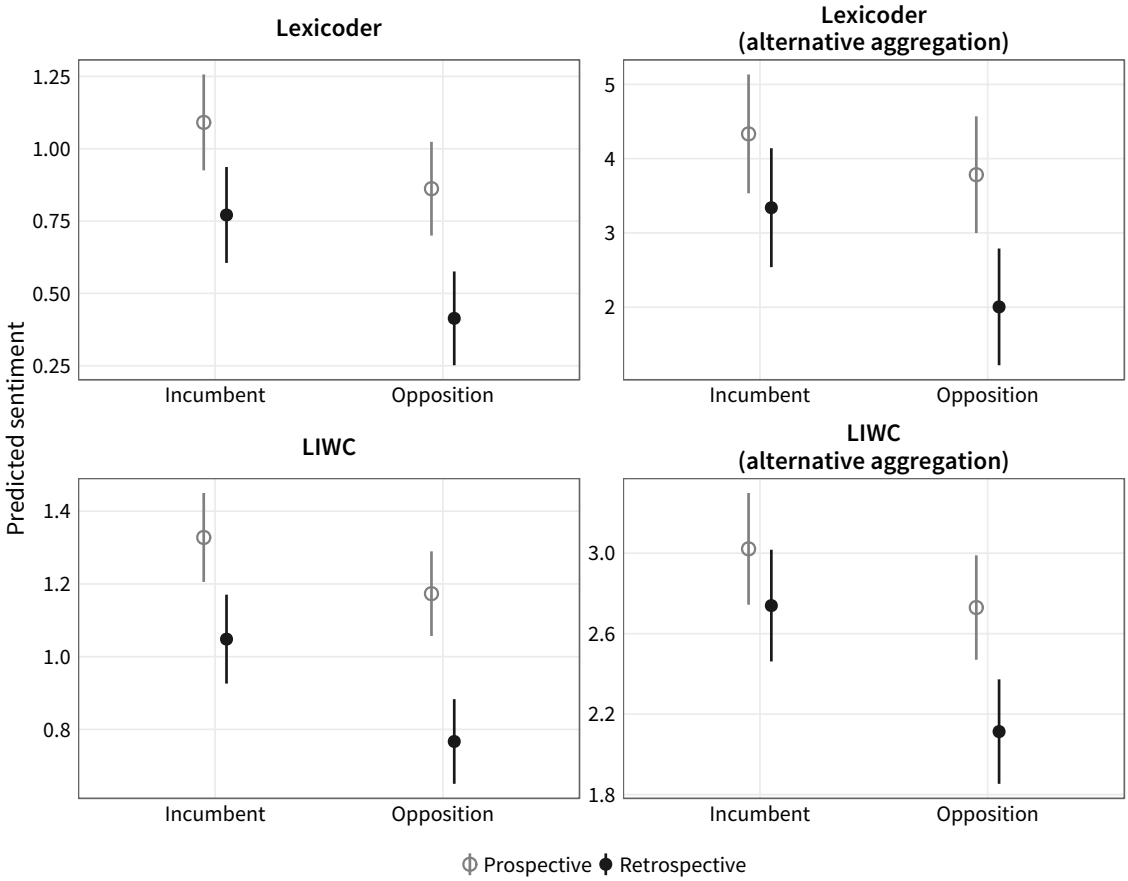
Figure B.6 plots the predicted values of Sentiment (measured with the Lexicoder Sentiment Dictionary and Proksch et al.'s (2019) German translation of this dictionary) conditional on left-right ideology. Figures B.7 and B.8 underscore the similarity of the results for incumbency status depending on the dictionary and aggregation of sentiment. Figure B.9 runs jackknife-style regressions by excluding one country at a time, running the model without this country, and plotting the predicted values. No single country influences differences between incumbents and opposition parties in prospective and retrospective statements.

Figure B.6: The impact of left-right positions on sentiment



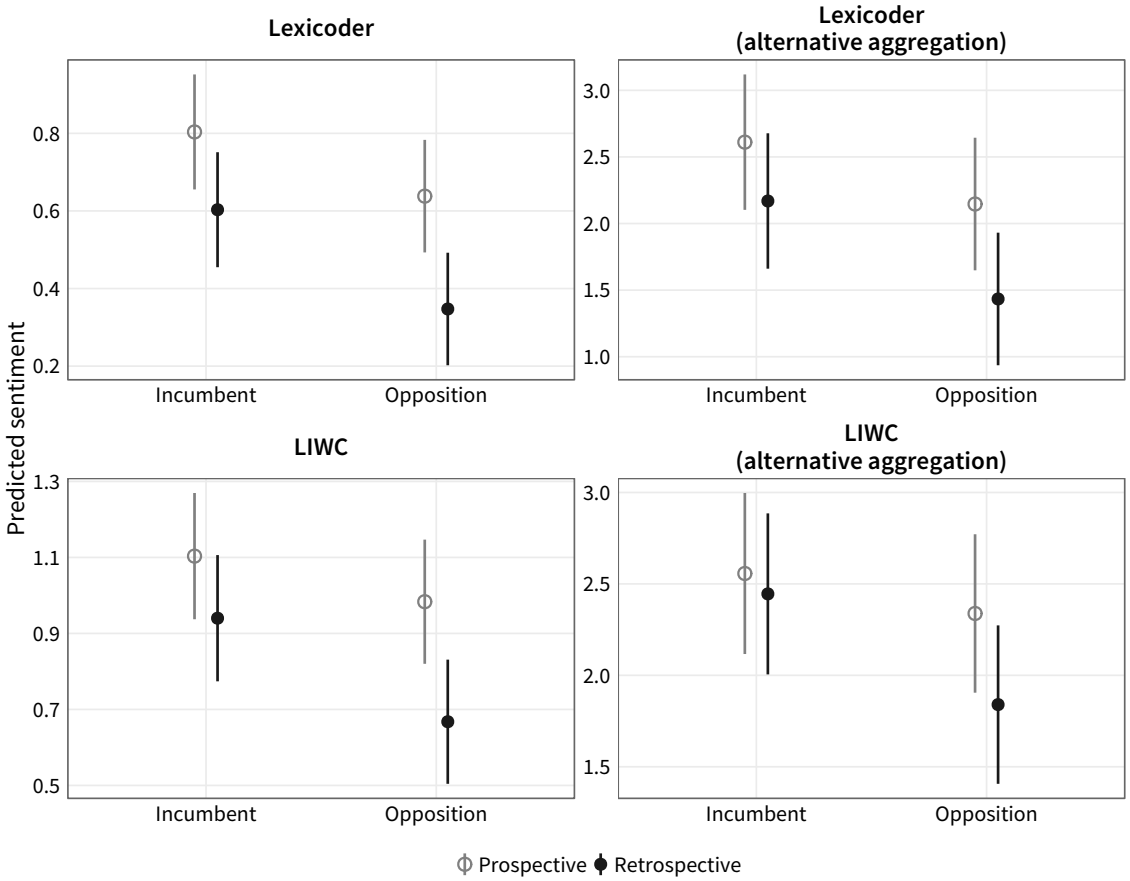
Note: Predicted values are based on Models 1 and 3 from Table 3.3. Shaded areas show 95 per cent confidence intervals.

Figure B.7: Predicting sentiment in prospective and retrospective manifesto sections with different measurements of sentiment (national elections)



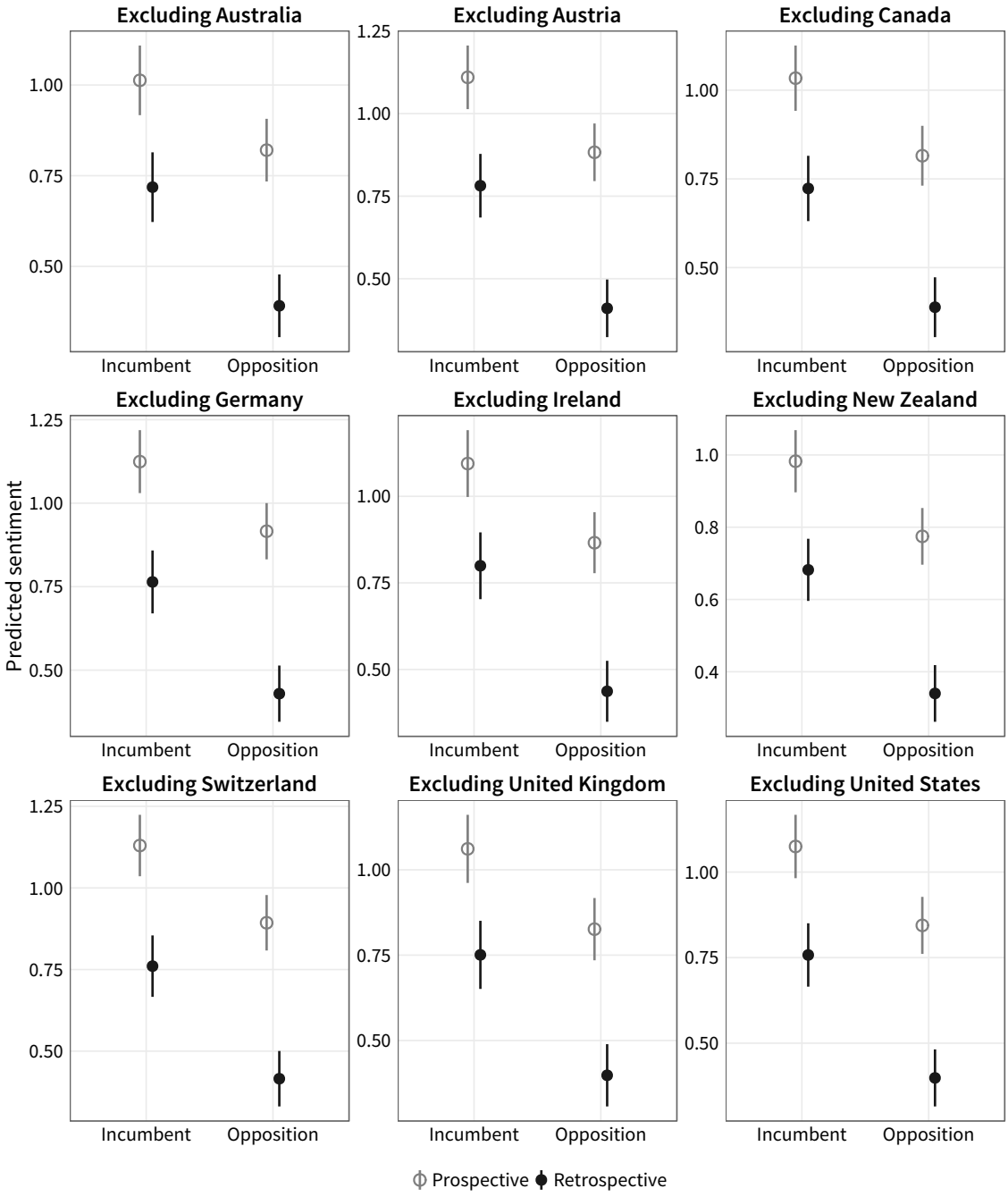
Note: Predicted values are based on Table B.8. Vertical bars indicate 95 per cent confidence intervals.

Figure B.8: Predicting sentiment in prospective and retrospective manifesto sections with different measurements of sentiment (German regional elections)



Note: Predicted values are based on Table B.9. Vertical bars indicate 95 per cent confidence intervals.

Figure B.9: Testing whether a specific country drives differences in sentiment



Note: The plot is based on nine separate regressions. Each regression reproduces Model 1 of Table 3.3, but excludes one of the countries in a jackknife regression style and runs the model on the eight remaining countries. Vertical bars indicate 95 per cent confidence intervals.

Table B.8: Predicting sentiment in party manifestos from national elections with different dictionaries and aggregations

	Model 1	Model 2	Model 3	Model 4
Opposition	-0.23^{***}	-0.15^{***}	-0.55^{***}	-0.29^{**}
	(0.03)	(0.04)	(0.14)	(0.09)
Class: Retrospective	-0.32^{***}	-0.28^{***}	-0.99^{***}	-0.28^{***}
	(0.03)	(0.03)	(0.11)	(0.07)
Seat share	0.11	0.00	0.41	0.02
	(0.12)	(0.13)	(0.55)	(0.34)
Left-right (RILE)	-1.83^{**}	-2.32^{***}	-6.33^{**}	-3.14^{*}
	(0.56)	(0.62)	(2.41)	(1.58)
Left-right^2 (RILE)	-0.95^{*}	-0.56	-4.58^{*}	-0.57
	(0.44)	(0.50)	(1.89)	(1.28)
Year	0.00	0.00^{**}	0.01	0.01^{***}
	(0.00)	(0.00)	(0.00)	(0.00)
Opposition × Class: Retrospective	-0.13^{***}	-0.13^{***}	-0.79^{***}	-0.33^{***}
	(0.03)	(0.03)	(0.14)	(0.09)
AIC	822.51	1026.29	4078.13	3148.91
BIC	887.99	1091.77	4143.62	3214.40
Log Likelihood	-398.25	-500.14	-2026.07	-1561.46
N	1138	1138	1138	1138
Num. groups: Manifesto	569	569	569	569
Num. groups: Election	144	144	144	144
Num. groups: Country	97	97	97	97
Num. groups: Party	9	9	9	9

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Note: Multilevel linear regressions with the sentiment (based on four different dictionaries) as the dependent variable. Model 1 uses the Lexicoder Sentiment Dictionary, Model 2 uses the LIWC dictionary instead. Models 3 and 4 reproduce Models 1 and 2, but use the calculation of sentiment used in Soroka (2012). Standard errors in parentheses.

Table B.9: Predicting sentiment in party manifestos from German Länder elections with different dictionaries and aggregations

	Model 1	Model 2	Model 3	Model 4
Opposition	-0.17^{***}	-0.12^{***}	-0.46^{***}	-0.22^{**}
	(0.03)	(0.03)	(0.09)	(0.07)
Class: Retrospective	-0.20^{***}	-0.16^{***}	-0.44^{***}	-0.11^{*}
	(0.02)	(0.02)	(0.05)	(0.05)
Seat share	0.34^{**}	0.30^{**}	1.28^{**}	0.68^{*}
	(0.12)	(0.12)	(0.40)	(0.30)
Left-right (Wordfish)	0.85^{*}	-0.28	3.14^{*}	-0.37
	(0.42)	(0.39)	(1.41)	(1.05)
Left-right^2 (Wordfish)	-1.66^{***}	-0.56	-6.78^{***}	-2.38^{**}
	(0.36)	(0.34)	(1.23)	(0.90)
Year	0.01^{***}	0.02^{***}	0.02^{***}	0.04^{***}
	(0.00)	(0.00)	(0.00)	(0.00)
Opposition × Class: Retrospective	-0.09^{***}	-0.15^{***}	-0.27^{***}	-0.39^{***}
	(0.02)	(0.03)	(0.07)	(0.07)
AIC	74.72	346.87	3240.63	2814.13
BIC	141.81	413.96	3307.72	2881.23
Log Likelihood	-24.36	-160.43	-1607.32	-1394.07
N	1288	1288	1288	1288
Num. groups: Manifesto	644	644	644	644
Num. groups: Election	141	141	141	141
Num. groups: Bundesland	16	16	16	16
Num. groups: Party	15	15	15	15

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Note: Multilevel linear regressions with the sentiment (based on four different dictionaries) as the dependent variable. Model 1 uses the Lexicoder Sentiment Dictionary, Model 2 uses the LIWC dictionary instead. Models 3 and 4 reproduce Models 1 and 2, but use the calculation of sentiment from Soroka (2012). Standard errors in parentheses.

Table B.10: Predicting sentiment on the level of sentences in party manifestos

	M1 (National)	M2 (Subnational)
Opposition	-0.09^{***}	-0.07^{***}
	(0.01)	(0.01)
Class: Retrospective	-0.22^{***}	-0.11^{***}
	(0.01)	(0.00)
Seat share	0.09	0.21^{***}
	(0.06)	(0.05)
Left-right	-13.11^{**}	7.70^{*}
	(5.05)	(3.73)
Left-right^2	-8.76^{*}	-22.07^{***}
	(3.56)	(3.74)
Year	0.00	0.00^{***}
	(0.00)	(0.00)
Opposition × Class: Retrospective	-0.10^{***}	-0.02^{***}
	(0.01)	(0.01)
AIC	849667.55	1542950.23
BIC	849805.38	1543096.66
Log Likelihood	-424820.78	-771462.11
N	297135	576187
Num. groups: Manifesto	501	644
Num. groups: Election	132	141
Num. groups: Party	88	15
Num. groups: Country	9	
Num. groups: Bundesland		16

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Note: Multilevel linear regressions with the sentiment on the sentence-level (based on the Lexicoder Sentiment Dictionary and Proksch et al.'s (2019) formula) as the dependent variable. Model 1 runs the model for manifestos from national elections, Model 2 uses the German regional election manifestos. Standard errors in parentheses.

Table B.11: Predicting the proportion of prospective statements in opposition parties' manifestos

	M1 (National)	M2 (Subnational)
Seat share	-0.43^{***}	0.02
	(0.13)	(0.08)
Left-right	-0.26^{*}	-0.28^{***}
	(0.11)	(0.07)
Left-right^2	0.12^{*}	0.12^{**}
	(0.05)	(0.04)
Year	0.00	0.00^{***}
	(0.00)	(0.00)
Number of words in manifesto (log)	0.02^{**}	0.00
	(0.01)	(0.00)
AIC	-534.98	-981.97
BIC	-496.52	-942.01
Log Likelihood	277.49	500.99
N	346	402
Num. groups: Election	141	139
Num. groups: Party	91	14
Num. groups: Country	9	
Num. groups: Bundesland		16

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Note: Multilevel linear regressions with the sentiment as the dependent variable. Model 1 focuses on opposition parties in national elections, Model 2 focuses on opposition parties in German regional elections.

Table B.12: Predicting absolute differences in prospective and retrospective sentiment

	M1 (Nat.)	M2 (Nat.)	M3 (Subn.)	M4 (Subn.)
Opposition	0.06*	0.08*	0.09***	0.09***
	(0.03)	(0.03)	(0.02)	(0.02)
Seat share		0.16		−0.00
		(0.10)		(0.10)
Left-right		−1.33***		0.39
		(0.37)		(0.27)
Left-right^2		−0.39		0.13
		(0.34)		(0.25)
Year		−0.00		−0.00
		(0.00)		(0.00)
AIC	410.91	415.50	-13.17	8.89
BIC	436.97	458.94	13.64	53.57
Log Likelihood	-199.45	-197.75	12.59	5.55
N	569	569	644	644
Num. groups: Election	144	144	141	141
Num. groups: Party	97	97	15	15
Num. groups: Country	9	9		
Num. groups: Bundesland			16	16

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Note: Multilevel linear regressions with the difference in sentiment in prospective and retrospective sections in each manifesto as the dependent variable. Model 1 focuses on the manifestos from national elections. Model 2 reproduces the regression for German regional elections. Standard errors in parentheses.

B.6 Performance and Validation of the Naïve Bayes Classification

To test the performance of the Naïve Bayes algorithm, I used held-out test sets, annotated by five crowd workers. The German and English test sets consist of around 300 sentences randomly sampled sentences. The following tables show the classification performance based on the training sets used throughout the paper. The column 'Execution time' reports how long it took to transform the raw data to a document-feature matrix and to classify each of the 300 sentences. Precision is measured as $\frac{TP}{TP+FP}$, where TP are the number of 'true positives' and FP 'false positives'. Recall divides the 'false positives" by the sum of 'true positives' and 'false negatives' ($\frac{TP}{TP+FN}$). The F1 score is a harmonic mean of precision and recall ($2 \times \frac{Precision \times Recall}{Precision+Recall}$).

Tables B.13 and B.14 compare the performance of the classifiers. Overall, the Naïve Bayes classifier offers the best results for both languages. Alternative classification methods, such as Support Vector Machine (SVM), Maximum Entropy Modeling, Gradient Boosting, and Bagging, perform equally or worse. Importantly, the Naïve Bayes classifier does not consistently overestimate one category which would result in systematic measurement error. Given the large amount of textual data to be classified, the computational efficiency of Naïve Bayes is a further advantage.

Table B.13: Comparison of classifiers for English text

Algorithm	Precision	Recall	F1	Execution time (sec.)
Naive Bayes	0.85	0.90	0.88	0.23
SVM	0.90	0.93	0.91	1.32
Maxent	0.93	0.89	0.91	1.47
Boosting	0.79	0.95	0.86	10.93
Bagging	0.95	0.87	0.91	47.01

As noted in the paper, I also tested as to whether the F1 score decreases if I predict the classes of manifestos published in earlier times or documents from the

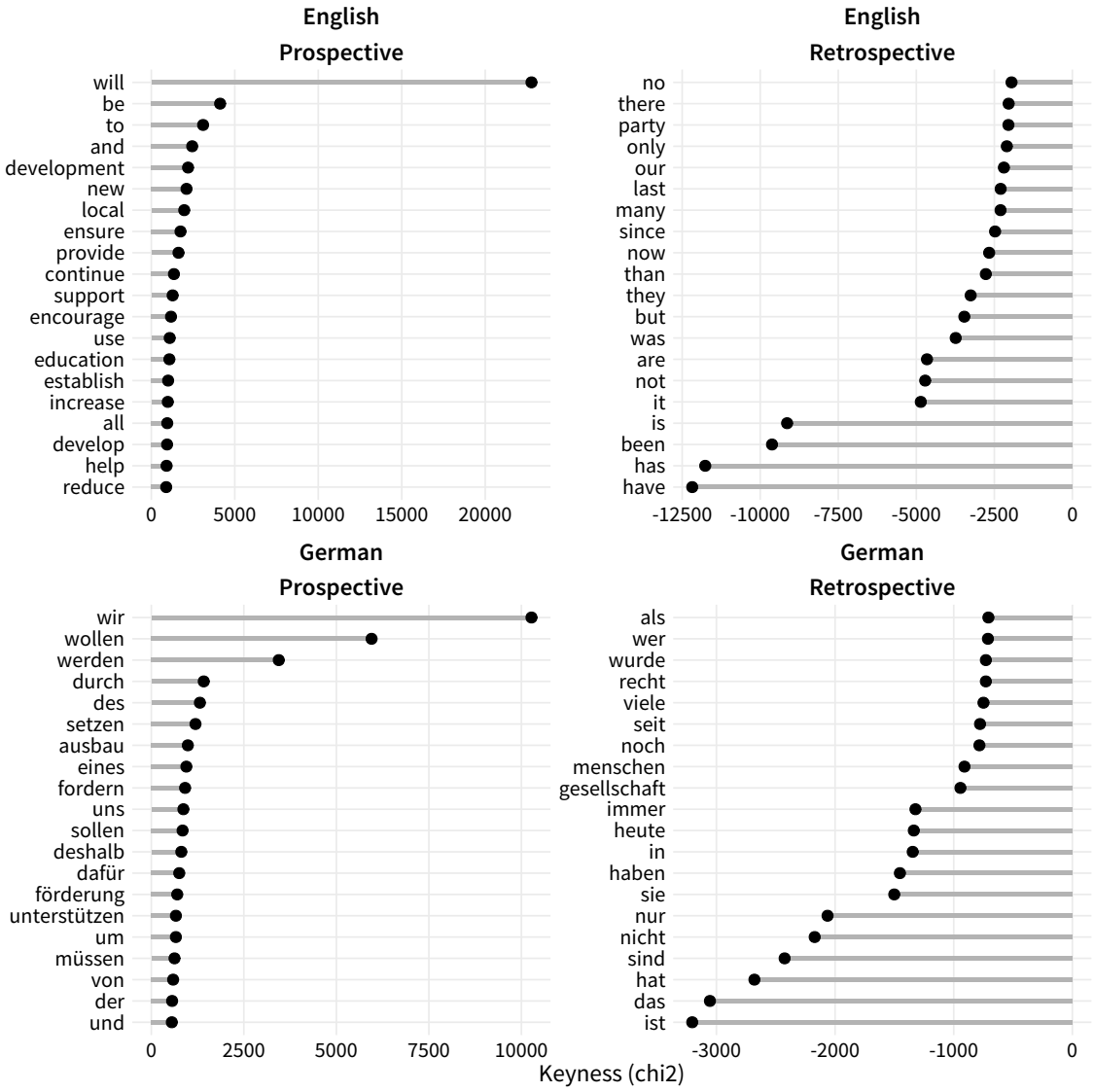
Table B.14: Comparison of classifiers for German text

Algorithm	Precision	Recall	F1	Execution time (sec.)
Naive Bayes	0.72	0.81	0.76	0.71
SVM	0.70	0.80	0.75	1.88
Maxent	0.66	0.83	0.73	3.90
Boosting	0.56	0.93	0.70	56.81
Bagging	0.60	0.95	0.74	270.80

same election as the training set. In an additional exercise I crowd coded 50–150 randomly sampled sentences from the party platform of the Conservative Party of Canada (2008), the manifesto of New Zealand's Labour Party from 1960, and the Irish Fianna Fáil manifesto from 2011. The F1 scores are remarkably stable over time and across countries (F1 ranging between 0.89 and 0.92 for the three manifesto samples).

Figure B.10 plots the so called 'keyness' of the classification. Keyness is a signed two-by-two association to identify frequent words in documents in a target and a reference group (Bondi and Scott 2010). The target group is "Prospective", the reference group are sentences classified as 'Retrospective'. The estimates are chi-squared values which have a positive sign if the observed value for a term exceeds the expected value. Words such as 'will', 'new' and 'ensure' have large positive values as these values occur much more frequently in sentences classified as 'prospective' than sentences classified as 'retrospective'. On the other hand, terms such as 'have', 'has', 'been' or 'is' have large negative values as they mostly occurred in sentences of the reference group 'Retrospective'. The terms from the 'keyness' statistics provide strong evidence for the face validity of the Naïve Bayes classifier.

Figure B.10: "Keyness" statistics for statements classified as retrospective and prospective



Note: The keyness statistics compare the differential associations of words with a target and a reference group. A positive (negative) value implies that a word is more more frequent than expected in the 'Prospective' ('Retrospective') class.

Appendix C

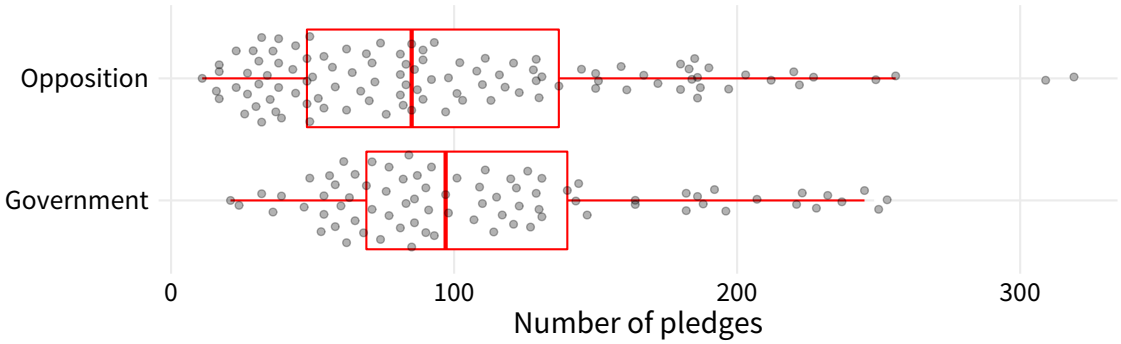
Media Coverage of Campaign Promises Throughout the Electoral Cycle (Supporting Information)

C.1 Descriptive Plots and Tables

Figure C.1 shows the number of pledges from the manifestos analysed in Thomson et al.'s (2017) comparative study of pledge fulfilment. Figure C.2 lists the pledges included in the studies by Thomson (2011), Thomson and Brandenburg (2018), Naurin and Oscarsson (2017), Belchior (2018), and Duval and Pétry (2018a), along with the proportions of respondents who evaluated the fulfilment incorrectly as well as the proportions of respondents who replied 'don't know'. Figure C.3 lists the effective number of parliamentary parties, Figure C.4 shows the number of government types within each country. Table 4.1 lists the number of promise-related sentences per cycle, the number of newspapers available for each cycle, and the average number of sentences per day (and per day and newspaper). Note that coverage, both in relative and absolute terms, is highest for newspapers from the United Kingdom. Over time,

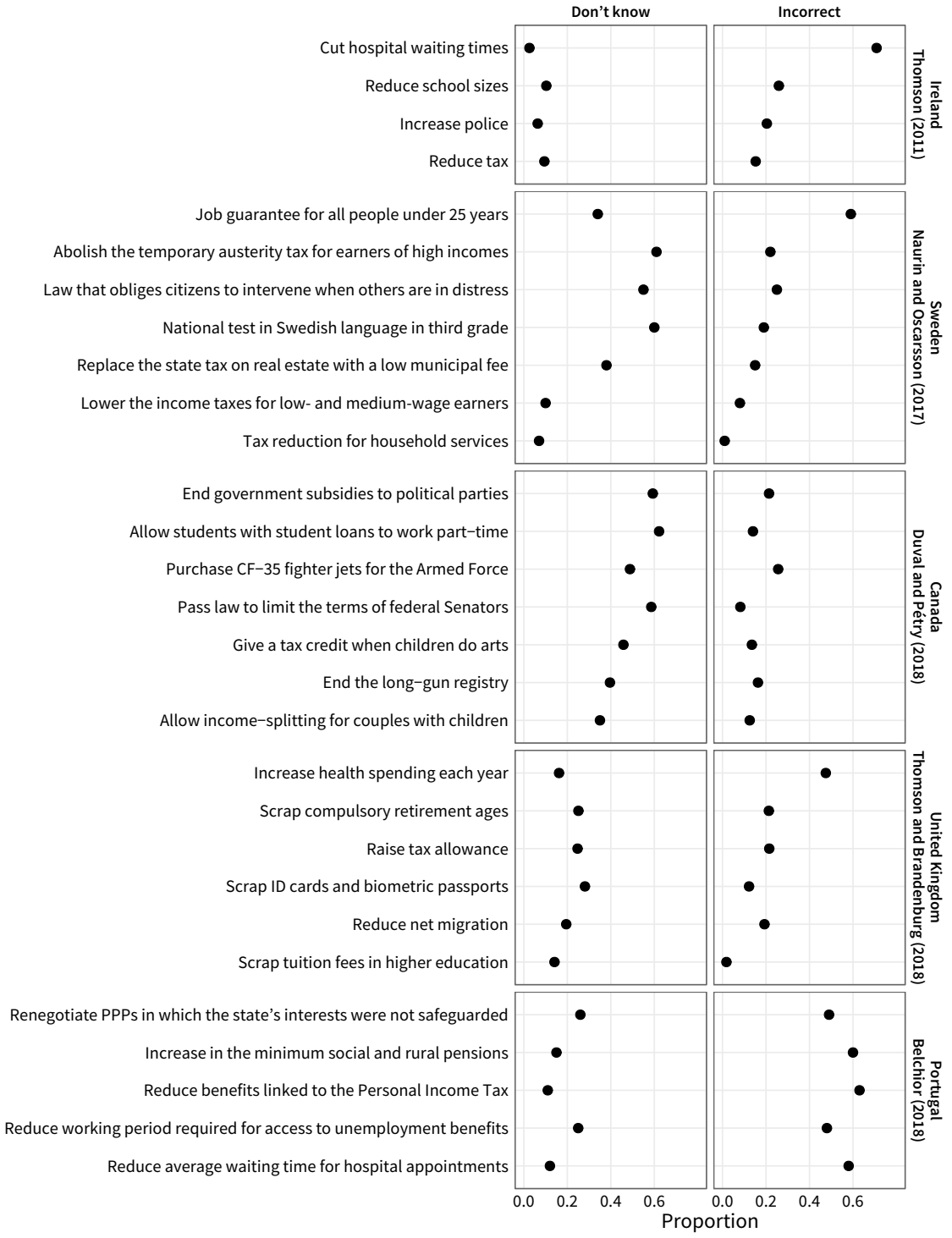
the coverage remains per day and outlet rather stable in relative terms, the only outlier is the cycle between 2015 and 2017 in the United Kingdom.

Figure C.1: The number of pledges in party manifestos, faceted by government status



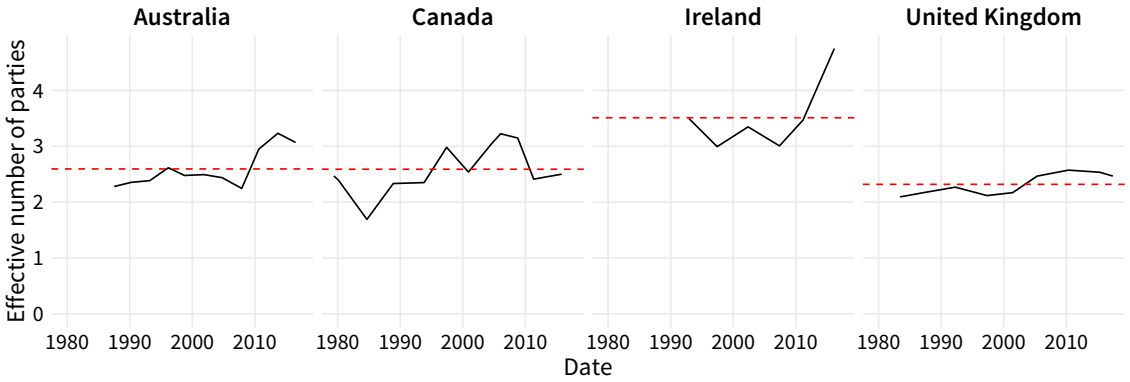
Note: Each dot marks one of the 190 manifestos analysed in Thomson et al. (2017). The boxplots show the median and interquartile ranges.

Figure C.2: The proportions of respondents who did not know about the fulfilment of a pledge and who made an incorrect evaluation



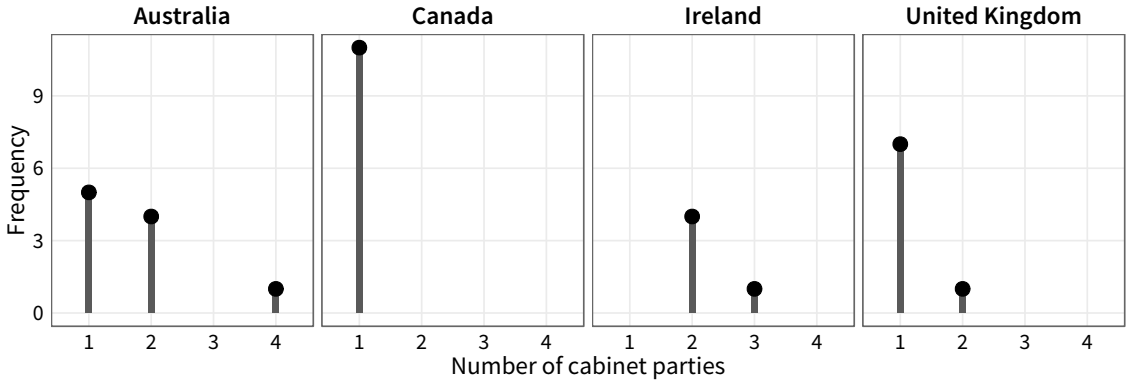
Note: Each dot shows one pledge from the four published studies on citizens' pledge evaluations (Thomson 2011; Thomson and Brandenburg 2018; Naurin and Oscarsson 2017; Belchior 2018; Duval and Pétry 2018a).

Figure C.3: Effective number of parliamentary parties (calculations based on the ParlGov dataset)



Note: The figure only considers the elections for which newspaper articles are available. Horizontal dashed lines show the average effective number of parliamentary parties for the periods under investigation.

Figure C.4: Number of government parties per cabinet (calculations based on the ParlGov dataset)



C.2 Classifying Pledge-Related Sentences

As described in the main part of the paper, newspapers with the highest circulation within each country were chosen. If possible, I also selected newspapers with different political leanings and different formats (broadsheet and tabloid). Articles are retrieved from NexisLexis. The search query followed the same pattern across all countries. The string retrieves articles that contain a term indicating a promise/pledge, or mentioning a the word party manifesto *and* contain the name of one of the main parties in the respective country.[1] Note that NexisLexis automatically finds singular, regular plural and possessive endings for search words. NexisLexis sometimes posts two or more versions of the same article. I removed these duplicated articles using pattern matching of the outlet, the title of the article, and the date. I only kept the version with a highest number of tokens.

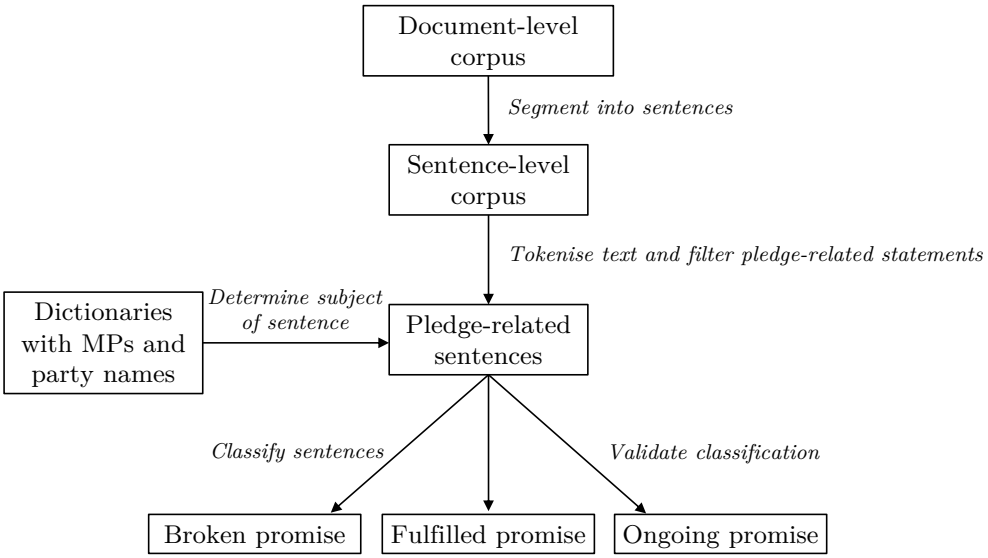
Figure C.5 outlines the classification. First, I load all newspaper articles into a text corpus and reshape the document-level corpus to the level of sentences using the `quanteda` R package (Benoit et al. 2018). The reshaped corpus consists of over 16.6 million sentences.[2] The main interest, however, lies in sentences that directly relate to promises, not all sentences from the articles. Therefore, I only keep sentences that contain words starting with `pledge*` or `promise*`. This reduces the text corpus by over 97 per cent to 486,718 sentences from 329,070 unique articles (for a similar classification method see, e.g., Soroka and Wlezien 2018).

Afterwards, I check whether the statement (or its contextual unit of ± one sentence) mentions a party or politician. I tokenize the pledge-related sentences and

[1] For example, the following string has been used for the UK newspapers: "`promise* OR pledge* OR manifesto* AND Labour* OR Conservatives* OR Tories* OR Conservative Party* OR Liberal Democrats* OR Liberal Democratic Party* OR UKIP* OR Scottish National Party* OR Democratic Unionist Party* OR DUP* OR SNP* OR Green Party* OR Plaid Cymru*`".

[2] Beside the duplicates, this corpus also includes versions of articles that have been published on the newspapers' websites. Because the offline and online news are almost identical, I focus only on printed newspapers.

Figure C.5: Retrieving, segmenting, and classifiying newspaper articles



compound party names as well as first and surnames of MPs. I apply a dictionary with party names and abbreviations and a second dictionary containing the names of politicians serving in each cycle. The dictionary was created using datasets from the website *EveryPolitician*[3]. For missing cycles I scraped the names and party affiliations from the websites of the national parliaments. This new dictionary consists of over 11,000 observations, where each observation represents one politician serving during one of the cycles. Note that "glob"-style wildcard matching is used to detect genitives and plural forms. In order to avoid too many false-positives, the names/parties are matched case-sensitive. For instance, the algorithm will classify `Labour_Party`, `Labour_Party's`, and `Labour` as a party, but not `labour`. I run the dictionary analysis of politicians in a loop for each cycle to ensure that I only count those politicians that served in the respective electoral cycle. Afterwards, I apply the dictionary with keywords for broken and fulfilled promises. If the number of terms from the 'broken' ('fulfilled') class exceeds the 'fulfilled' ('broken') keywords, the sentence is coded as 'broken' ('fulfilled'). Only 270 sentences (less than 0.05 per cent of the corpus of relevant sentences) contain an equal number of terms from the

[3]https://everypolitician.org.

'fulfilled' or 'broken' classes. Results do not change when adding these cases to either 'broken' or 'fulfilled'.

C.3 Validation

Validation is essential for all methods that use text-as-data approaches (Grimmer and Stewart 2013). I validate the classification in three ways. First, I employ an alternative search query that looks up articles based on policies, not pledge-related terms. Second, I compare the automated classification to crowd coders. Third, I check face validity of the classification.

C.3.1 Selection of Newspaper Articles

I downloaded articles based on pledge-related keywords instead of actual pledges from manifestos. The question remains as to whether this search query systematically excludes articles that mention the progress or fulfilment of a campaign promise *without* using terms starting with `pledge*/promise*`. Therefore, I also chose a different procedure for retrieving articles. I focused on the six salient pledges that were evaluated by respondents in the 2015 British National Election Study (Thomson and Brandenburg 2018). For the three months prior to the 2015 General elections, I retrieved all articles mentioning one or more of the pledges. I decided to employ a very broad search query to filter articles on these policies published in Daily Mail, the Daily Mirror, the Daily Start, the Financial Times, the Independent, the Mail on Sunday, The Guardian, The Sun, and The Times. Table C.1 lists the keywords used for each pledge. In a second step, I checked for each article whether it mentions at least one British political party to ensure that the article is about British politics. Recall that this NexisLexis search query does *not* filter based on pledge-related terms. Thus, this separate search query allows to check whether articles on concrete and salient policies indeed contain pledge-related terms, and whether the keyword search based on `promise*/pledge` could extract some policies systematically.

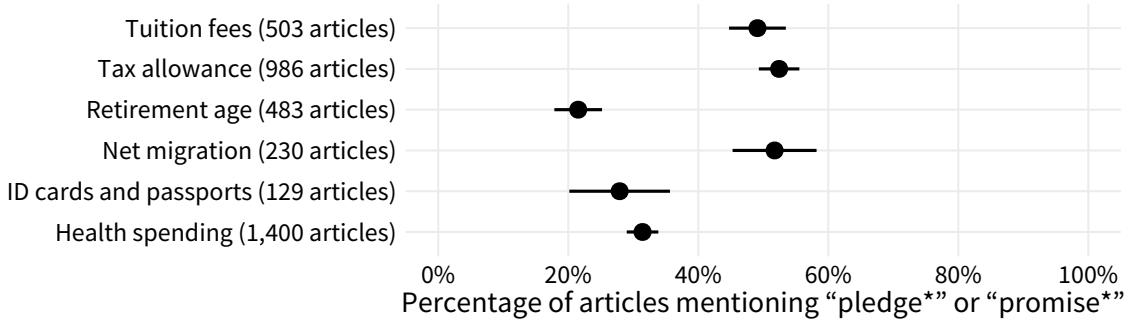
Overall, I retrieved 3,726 articles that contain at least one of the keywords for the testable pledges and a political party. Figure C.6 plots the proportion of articles

Table C.1: Search query for retrieving articles about pledges mentioned in Thomson and Brandenburg (2018)

Pledge	Query
Scrap University tuition fees	`tuition* OR fees*`
Raise the tax-free personal allowance to £10,000	`tax* OR allowance*`
Scrap compulsory retirement ages	`retire* OR age*`
Reduce annual net migration to less than 100,000	`migrat*`
Scrap ID cards and biometric passports	`id OR card* OR biometric passport*`
Increase health spending	`health*`

that contain pledge-related terms for each of the policies. With a range between 22 and 52 per cent, on average, 38 per cent of the news articles include at least one word indicating a promise (either `promise*` or `pledge*`). This robustness check suggests that text corpus used in this paper does not systematically exclude relevant reports about specific promises. Therefore, it seems reasonable to look up relevant articles based on pledge-related keywords instead of retrieving newspaper articles based on terms relating to concrete policies.

Figure C.6: Articles about policies used in Thomson and Brandenburg (2018) that include pledge-related terms

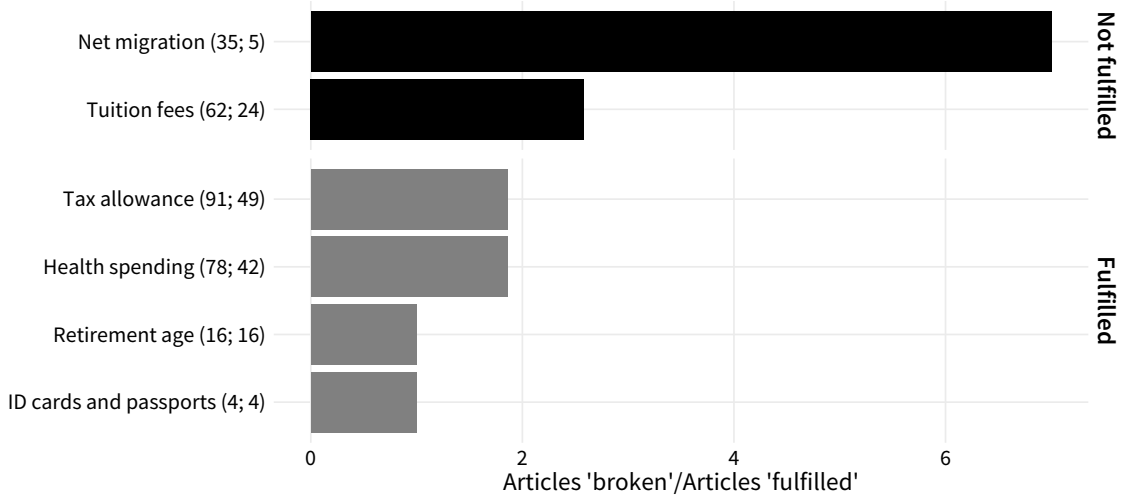


Note: The plot shows the proportion of articles from three UK newspapers published in the three months prior to the 2015 General Election, along with 95 per cent confidence intervals.

C.3.2 Crowd-Sourced Coding

To test whether the classification into broken, fulfilled, and ongoing promises is meaningful, I compare various text-as-data methods with aggregated crowd-sourced codings. To assess the accuracy of various classification methods, I compare the

Figure C.7: The ratio of articles classified as mentioning the breaking and fulfilment of a promise

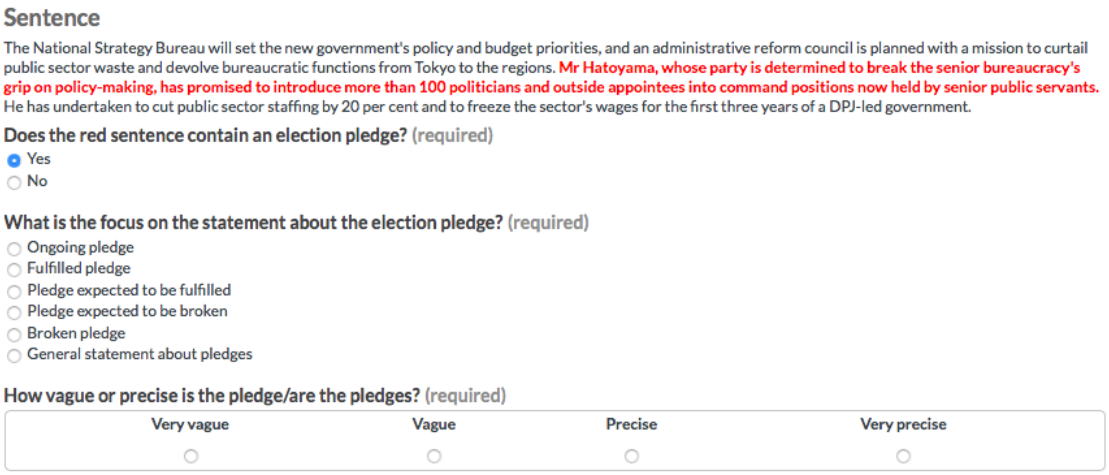


Note: The first number in brackets shows the absolute number of articles classified as 'broken', the second number indicates the number of articles classified as 'fulfilled'.

crowd-sourced coding of a random sample of 400 sentences to the performance of various classifiers on the same set of sentences. Recently, crowd-sourced text coding has been employed in many studies (Benoit et al. 2016). I recruit workers through the platform Figure Eight (previously named CrowdFlower). The workers interested in the coding job receive detailed instructions and several examples on how to code sentences (Section C.4). An entry test ensures that workers understand the coding instructions. Crowd workers who pass the entry test evaluate statements based on the coding instructions. To remove 'spammers' from the coding job, 20 per cent of all sentences are test questions with a pre-defined answer key. If respondents do not answer more than 80 per cent of these test questions correctly, they get removed from the job and their previous codings are not considered in the exercise. Each sentence is coded by three workers and aggregated using the most frequent answer. Figure C.8 shows a screenshot of the coding job. Workers are asked to judge the class of the sentence highlighted in red.

Table C.2 compares the dictionary-based classification with the most frequent crowd-sourced coding of the same set of sentences. I assigned the 'expected to be

Figure C.8: Screenshot from crowd coding job



fulfilled'/'expected to be broken' classes to 'fulfilled'/'broken'. The classification does not change when excluding these statements. The average F1 score across the three classes amounts to 0.63 (with a range from 0.61 to 0.7). The F1 score for the class 'fulfilled' is the lowest one, since the dictionary picked up many 'false positives', which results in low precision (Table C.3). Yet, this difference is not a major concern for the substantive interpretation of the results. The higher number of reports on broken than on fulfilled promises – measured through the dictionary – results in an underestimation rather than an overestimation of the 'true' difference.

Table C.2: Classification of dictionary-based classifier

	Broken	Fulfilled	Other
Predicted: Broken	138	1	59
Predicted: Fulfilled	13	23	20
Predicted: Other	43	4	103

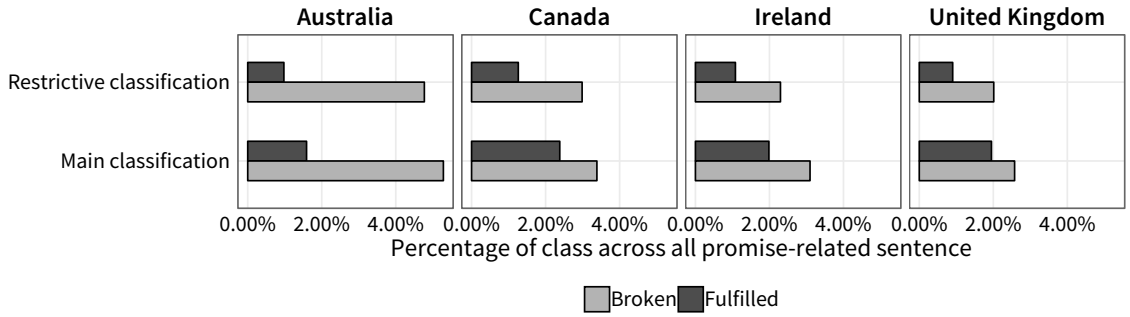
Table C.3: Performance of dictionary-based classification

	Class	F1 score	Precision	Recall
1	Class: Broken	0.70	0.69	0.70
2	Class: Fulfilled	0.61	0.49	0.80
3	Class: Other	0.63	0.68	0.58

The choice of keywords for the dictionaries could have large impact on the results. In order to test whether an even more restrictive classification reduces the differences

between statements 'broken' and 'fulfilled' promises, I also classify the sentences based on a very small set of terms: the dictionary for 'broken' only consists of `broke*, failed_to, not_fulfilled, and not_kept`, whereas the class 'fulfilled' consists of `fulfilled, kept, not_broken, and not_break`. As Figure C.9 indicates, the very restrictive classification even increases the differences between both classes.

Figure C.9: Comparing the percentage occurences of statements on 'broken' and 'fulfilled' based on two different classifications



C.4 Instructions for Crowd Coders

Overview

This task involves reading sentences from English newspapers and judging whether the sentence contains an election pledge (promise). An election pledge is a promise that commits a party or politician to one specific action or outcome. For the sentence highlighted in red, enter your best judgement. The coding task includes up to three questions about a statement.
First, you are asked whether a statement is about politics and deals with election pledges. Second, if the statement is about a pledge, you need to code the focus of the statement based on the following categories: fulfilled promise, promise expected to be fulfilled; broken promise; promise expected to be broken; ongoing promise; or a general statement about the nature of election promises. Third, you are asked to evaluate how precise the statement outlines one or more pledges. If you are not entirely sure about the context of the highlighted sentence, read the surrounding sentences as well. Ultimately, your judgement should focus on the sentence in red font. Below we define the categories and provide examples.
Your feedback is highly appreciated. Please take part in the contributor satisfaction survey. Thank you very much!

Examples

1) Does the statement refer to politics and election pledges?

Yes: Pledge

- "Most people remember past Liberal promises: their opposition to free trade; or their pledge to eliminate the Goods and Services Tax."
- "The promise now is to get back to the September 2015 level within 12 months and redeem the original pledge before the next election."

Explanation: The statements clearly relate to an election promise.

No pledge(s) mentioned

- "Political intervention will reduce confidence in the impartiality and consistency of the decisions."
- "The crowd waved signs including 'Promises made, promises kept', 'Lefty media lies' and 'Women for Trump'."

Explanation: Although the statements are broadly related to election pledges, they do not outline any specific promise made by a political party or a politician. The second statement is about a public protest, but does not include any promises made

by a party or politician, nor does it contain a journalist's judgement about election promises.

2) What is the focus of the statement about the election pledge?

Fulfilled pledge

- "The government fulfilled its promise on beginning with the demilitarisation announced after the negotiations last month."

Explanation: The statement clearly states that the government fulfilled a pledge on demilitarisation.

Pledge expected to be fulfilled

- "The Labour manifesto pledged to do away with the culture of secrecy in public life, and it seems that the government will keep this promise."

Explanation: While the statement does not unambiguously state that the promise has been fulfilled, it is expected or likely that the promise will be fulfilled.

Ongoing pledge

- "He also promised that it would achieve its target of cutting public grants for loss-making services from Pounds 1 billion in 1983 to Pounds 605 million by March 1989, one year earlier than promised."
- "Wilkie said Abbott should be 'a man of his word' and fulfil his election promise to let the party room decide on the issue of a free vote."

Explanation: While the statements deal with an election promise, the sentences do not contain any information on whether the promise has been fulfilled or not.

Pledge expected to be broken

- "Thus the promise of maintaining a 10 per cent GST is threatened only by the ALP's looming inability to satisfy the groups it has promised to appease."

Explanation: The statement assumes that a promise is very likely to be broken ("threatened by the inability to satisfy the groups it has promised to appease").

Broken pledge

- "'He's now broken his promise about Senate reform,' said Liberal Leader Michael Ignatieff, referring to Harper's pledge to allow voters - not the prime minister - to choose senators.
- "The Labour manifesto pledged to do away with the culture of secrecy in public life, but the draft bill fell well short of expectations and promises."

Explanation: The statements clearly highlight that the promises made by a party have not been fulfilled.

General statement about the nature of election pledges

- "We have been told by recent political leaders that there are such things as core and non-core promises and that only those promises given in writing count because verbal promises may change."

Explanation: The statement refers to election pledges in a very broad sense, but instead of offering information on any concrete policy, it contains a general normative judgement on the nature of election pledges.

3) How vague or precise is the pledge described in the sentence?

Very vague

- "The Labour manifesto pledged to do away with the culture of secrecy in public life, but the draft bill fell well short of expectations and promises."

Explanation: It is very difficult to assess what is meant by the 'culture of secrecy in public life'. This expression could imply a lot of policies.

Vague

- "I gather he thinks Mr Blair has broken his promises on education and we were better off with Margaret Thatcher."

Explanation: Although the statement mentions one or more promises on education, we receive no information about which promises on education the sentences refers to.

Precise

- "And the NDP promised tax breaks for families and small businesses, with some increases for corporations."

Explanation: The statement refers to tax breaks and increases for corporations (clearly a specific election promises), but the exact details are missing.

Very precise

- "An increase in the inheritance tax threshold on family homes at an estimated annual cost of £1 billion – one of the main pledges that the Tories have promised to deliver."

Explanation: The statement lists a very precise and testable election pledge including the concrete goals of the pledge.

C.5 Additional Plots and Regression Tables

Figure C.10 plots the monthly number of promise-related statements per country. Vertical dotted lines indicate the dates of elections. Figure C.11 plots the proportions of pledge-related statements of the three classes (ongoing, broken, fulfilled). Each dot marks one newspaper in one electoral cycle.

The models in Table C.4 are used to predict the coverage of each type of promises throughout the electoral cycle (see Figure 4.7). Model 1 focuses on the subset of sentences that has been classified as fulfilled promises, Model 2 focuses on the subset of broken promises, while Model 3 only uses sentences that mention a promise, but do not contain any information on the fulfilment or breaking of a promise. Figure C.13 reproduces Figure 4.5 after removing observations with very high values of the dependent variable (ratios exceeding 4).

Figure C.10: The montly count of promise-related statements across all newspapers

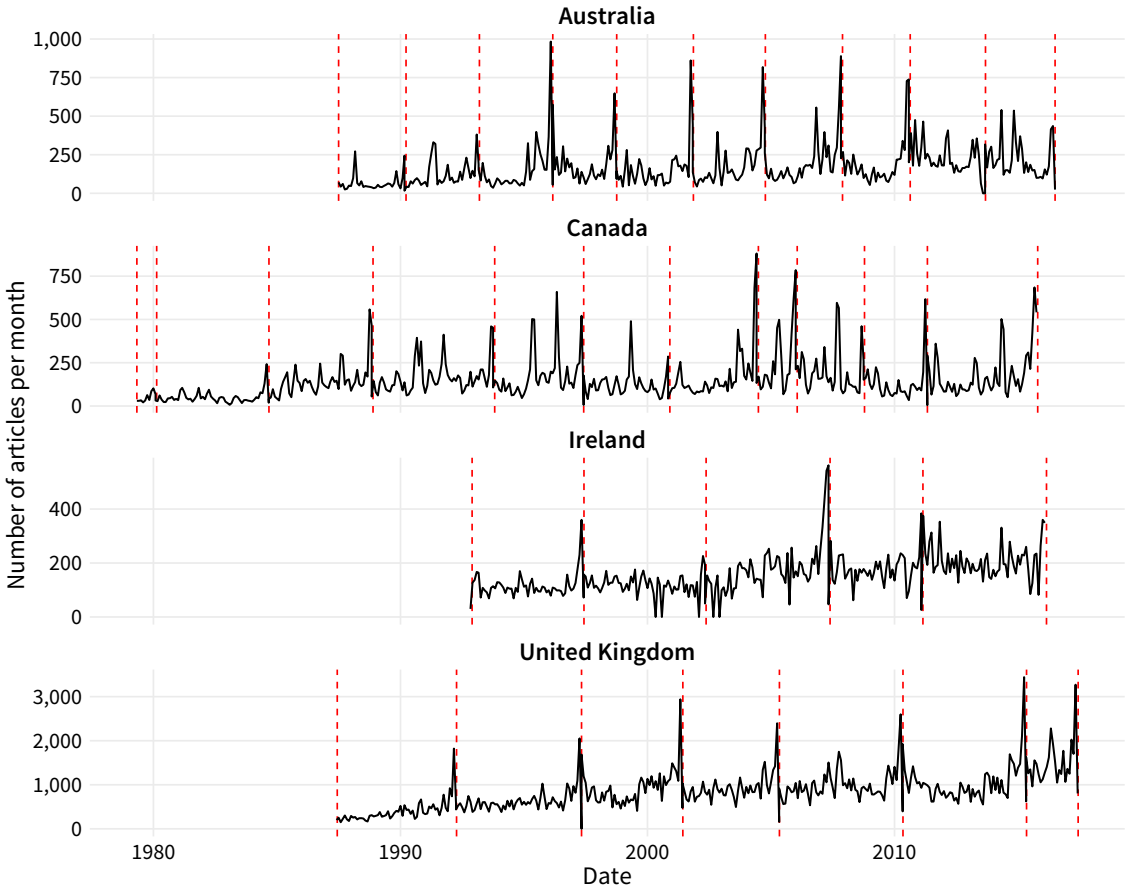
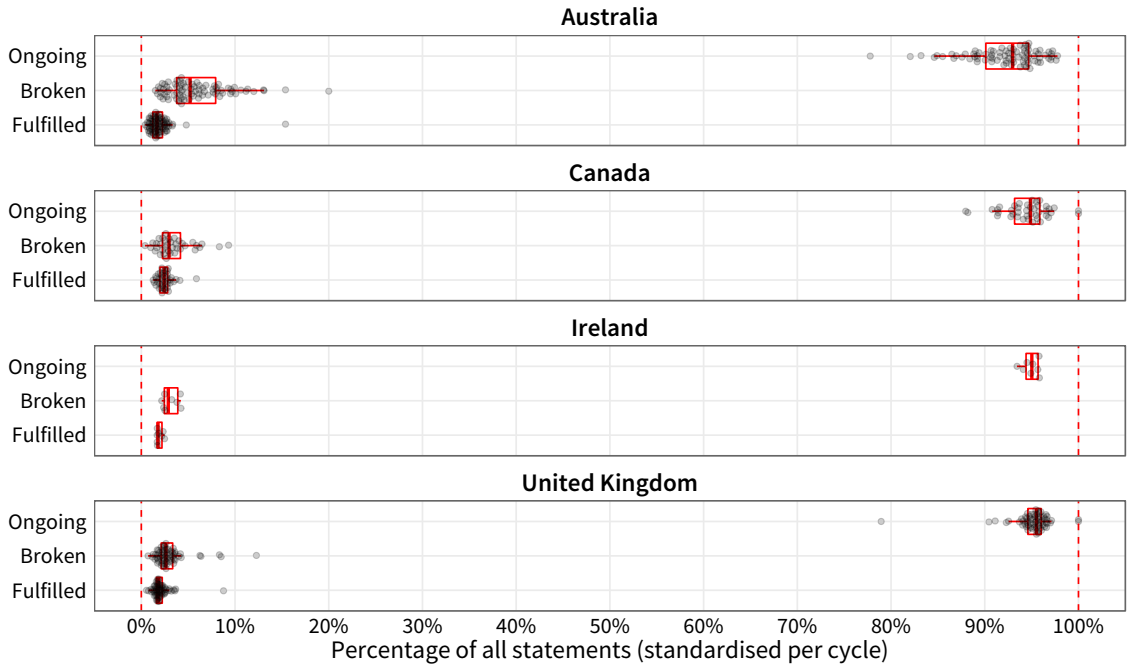
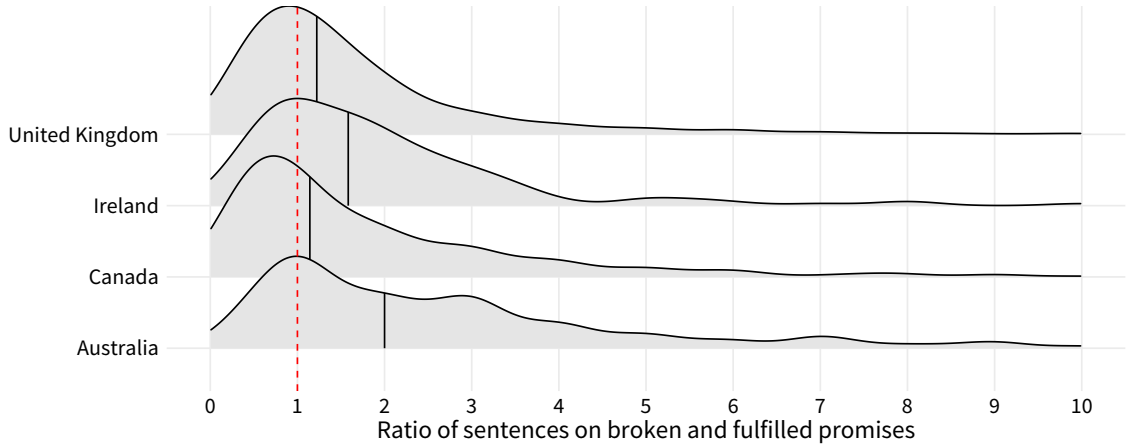


Figure C.11: The proportion of pledge statements across classes and countries



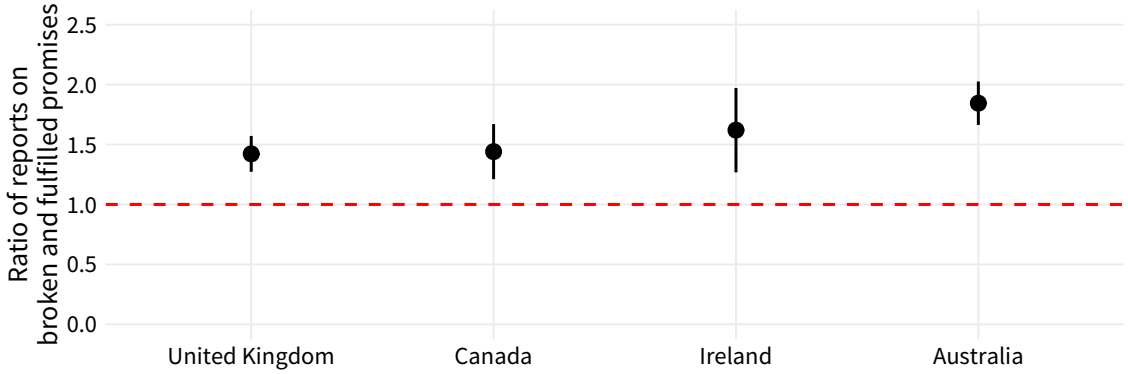
Note: Each dot marks the proportion of the respective class in one newspaper during one electoral cycle.

Figure C.12: Density of newspaper-quarter ratios of statements on broken and fulfilled promises



Note: A value larger than 1 means that the the coverage of broken promises exceeded the proportion of coverage of fulfilled promises. The vertical black lines show the median values for each country. The mean values are around twice as high for each country.

Figure C.13: Predicting the ratio of reports on broken and fulfilled promises per quarter for each country, excluding outliers



Note: The predicted values reproduce the model used to produce Figure 4.5, but 'outlier' observations with ratios exceeding 4 are removed before running the model. This approach reduces the sample to 67 per cent of the original dataset. A ratio of 1 implies that the number of statements on broken promises equals the number of statements on fulfilled promises. Vertical bars show 95 per cent confidence intervals.

Table C.4: Predicting the count of articles per day

	M1 (Fulfilled)	M2 (Broken)	M3 (Ongoing)
Intercept	-2.99^{***}	-2.41^{***}	0.67^{***}
	(0.15)	(0.11)	(0.03)
N Articles (lag)	0.14^{***}	0.02^{***}	0.23^{***}
	(0.01)	(0.00)	(0.00)
Electoral cycle	9.32^{***}	18.75^{***}	34.14^{***}
	(2.42)	(2.38)	(0.90)
Electoral cycle^2	19.12^{***}	7.90^{**}	26.30^{***}
	(2.46)	(2.43)	(0.93)
Electoral cycle^3	15.53^{***}	26.56^{***}	19.47^{***}
	(2.48)	(2.41)	(0.93)
GDP change (lagged by quarter)	-0.03^{*}	0.00	-0.01^{***}
	(0.01)	(0.01)	(0.00)
Month of cabinet change	0.52^{***}	0.14	0.39^{***}
	(0.07)	(0.07)	(0.03)
AIC	43398.43	57218.38	258484.35
BIC	43736.53	57556.46	258822.45
Log Likelihood	-21660.22	-28570.19	-129203.18
Deviance	22794.18	24703.57	50351.98
Num. obs.	43008	42994	43011

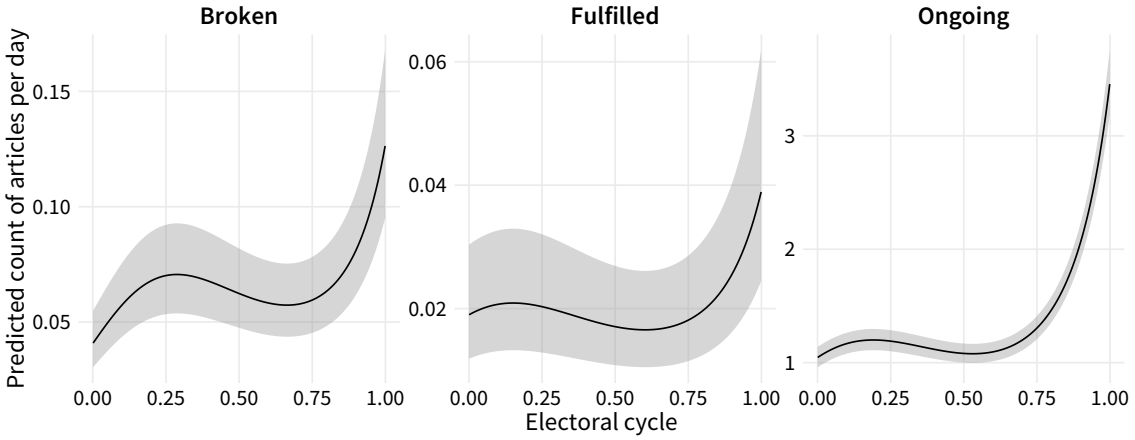
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Note: Negative binomial regressions with dummies for country-cycle observations (omitted from regression table). Standard errors in parentheses.

Table C.5: Number of relevant sentences for each category and country (subset of statements that mention an MP or a party)

Country	Ongoing	Broken	Fulfilled	Total
Australia	35,424 (92.8%)	2,199 (5.8%)	541 (1.4%)	38,164
Canada	26,816 (93.7%)	1,105 (3.9%)	698 (2.4%)	28,619
Ireland	16,849 (94.4%)	687 (3.8%)	321 (1.8%)	17,857
United Kingdom	192,604 (95.5%)	5,468 (2.7%)	3,701 (1.8%)	201,773

Figure C.14: Predicting the number of pledge-related articles per day that also mention a party or MP, separately for each class



Note: The predicted values are based on negative binomial regression models with the number of daily articles as the dependent variable. Note that the y-axes differ for each facet to show the developments within each class. The figures are based on Models 1–3 of Table C.6. Grey areas show 95 per cent confidence intervals.

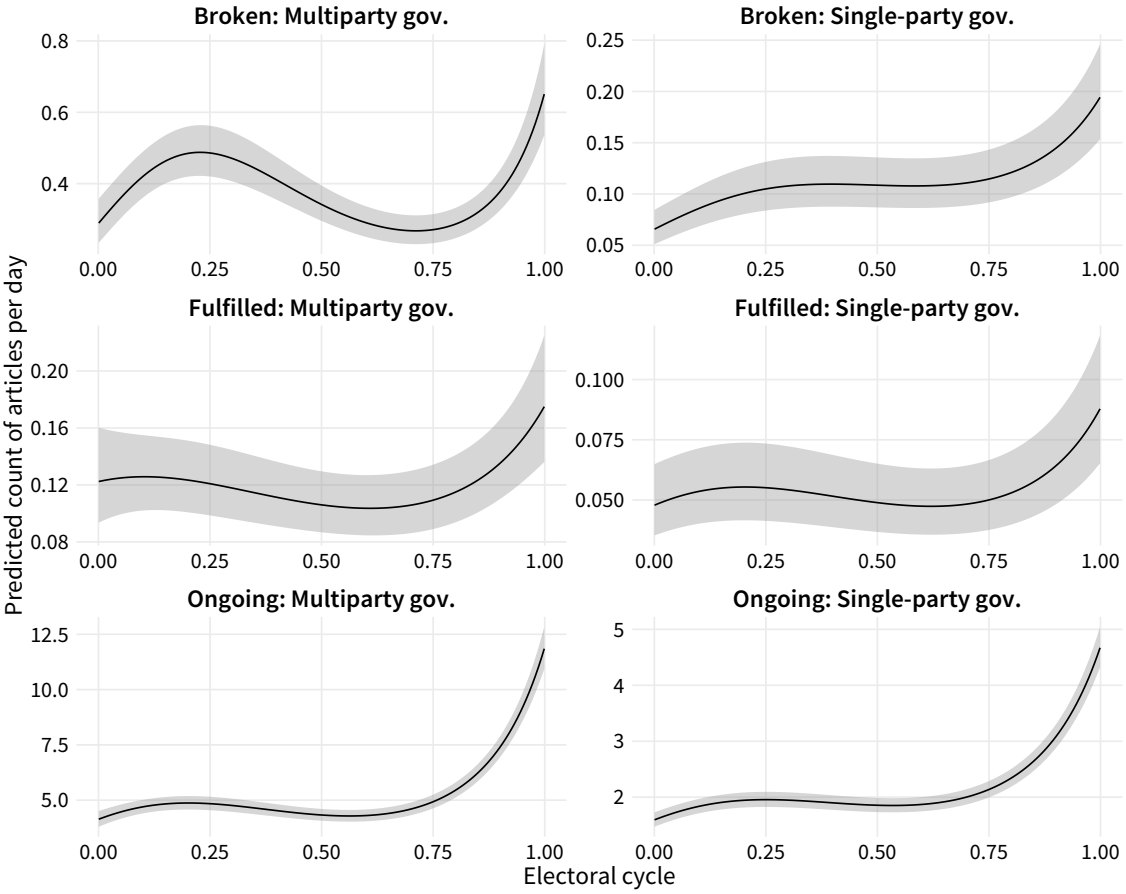
Table C.6: Predicting the count of articles per day for sentences that mention a political actor

	M1 (Fulfilled)	M2 (Broken)	M3 (Ongoing)
Intercept	-3.90^{***}	-2.87^{***}	0.26^{***}
	(0.23)	(0.14)	(0.04)
N Articles (lag)	0.01^{***}	0.02^{***}	0.05^{***}
	(0.00)	(0.00)	(0.00)
Electoral cycle	12.20^{***}	20.39^{***}	38.15^{***}
	(3.01)	(2.77)	(0.89)
Electoral cycle^2	27.74^{***}	8.45^{**}	33.55^{***}
	(3.06)	(2.84)	(0.91)
Electoral cycle^3	19.07^{***}	29.51^{***}	20.12^{***}
	(3.09)	(2.80)	(0.92)
GDP change (lagged by quarter)	-0.06^{**}	0.01	-0.02^{***}
	(0.02)	(0.01)	(0.00)
Month of cabinet change	0.71^{***}	0.17	0.50^{***}
	(0.08)	(0.09)	(0.03)
AIC	27028.80	40142.26	193916.03
BIC	27364.18	40477.73	194249.30
Log Likelihood	-13475.40	-20032.13	-96919.02
Deviance	15395.78	18539.89	42280.01
Num. obs.	40115	40214	38002

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.
Note: Negative binomial regressions with dummies for country-cycle observations (omitted from regression table). Standard errors in parentheses.

The sample consists of 11 multiparty cabinets and 21 single-party governments. Figure C.15 reproduces the regressions separately for single- and multiparty cabinets. The aggregated smoothed shapes for ongoing promises are almost the exact same, and the shapes for fulfilled promises are also similar. The only difference is more pronounced increase in coverage of broken promises at around 20 per cent of the cycle for multiparty cabinets. The requirement of forming a coalition usually results in compromises by all parties involved which decreases mandate fulfilment (Thomson et al. 2017). This temporary rise in coverage might be a consequence of compromises that result in broken promises for some parties. Yet, because the loess regressions are not always consistent with the predicted values from the negative binomial regressions and because the confidence intervals are wide, this trend should be treated with caution and further analysed in case studies.

Figure C.15: Predicting the number of articles per day for single-party and multiparty governments, separately for each class

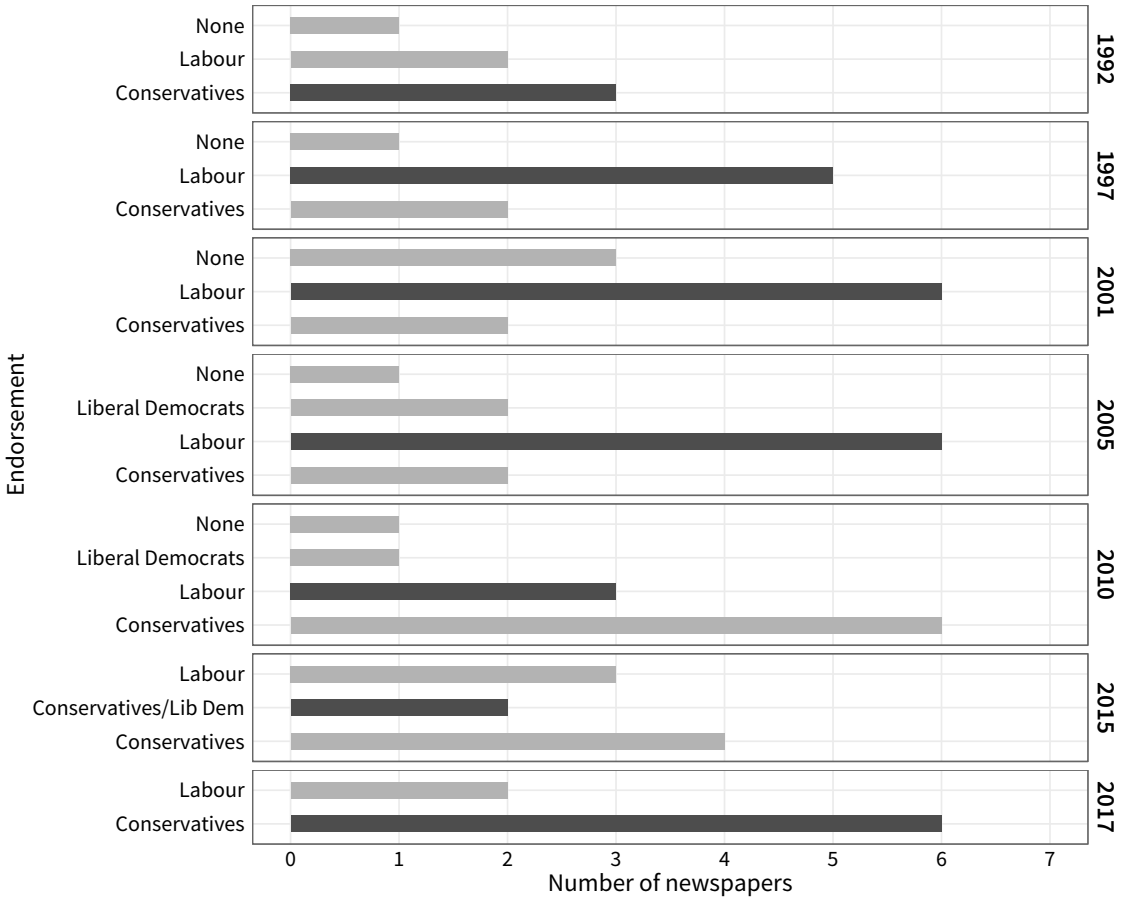


Note: The predicted values are based on negative binomial regression models with the number of daily articles as the dependent variable. Note that the y-axes differ for each facet to show the developments within each class. The figures are based on similar models as in Table C.4. However, I first subset the dataset for each government types, and run each of the three models (for statements about fulfilled, broken, and ongoing sentences) separately for single- and multiparty cabinets. The grey areas show 95 per cent confidence intervals.

C.6 Analysing Newspaper Polarisation in the United Kingdom

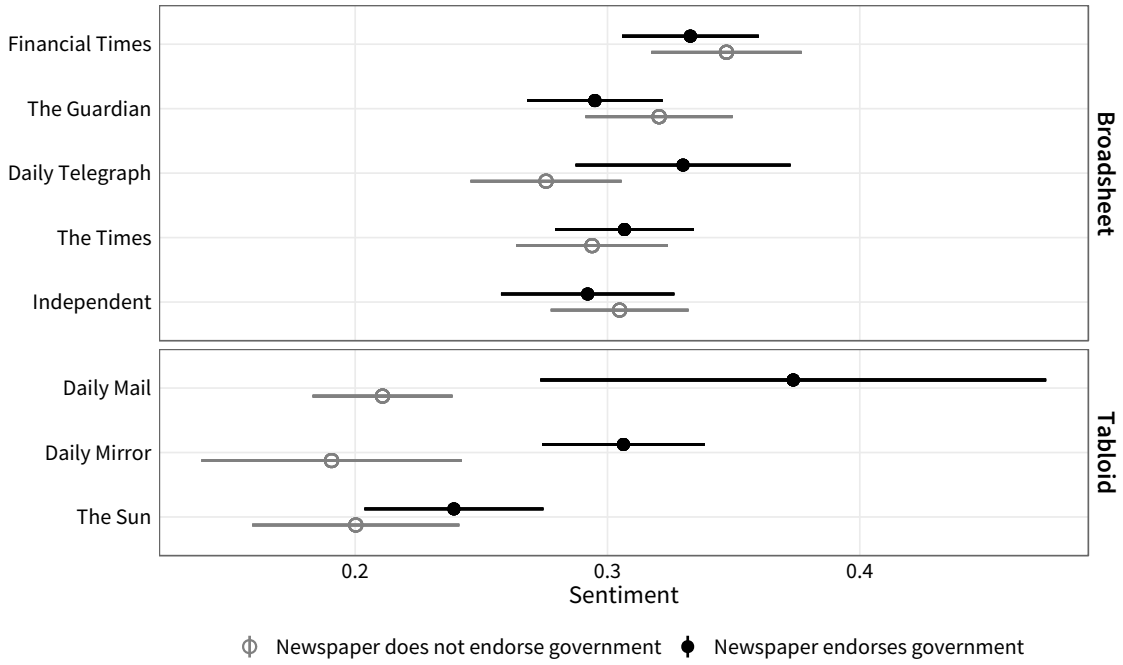
The polarised media landscape in the United Kingdom provides an excellent case for testing whether partisanship of news outlets affects the reporting on pledges when the endorsed party is (not) in government. I classify the partisanship of each newspaper and cycle based on the data provided in Wring and Deacon (2010, 2018). Figure C.16 shows how many newspapers endorsed one of the parties or remained neutral. The incumbent party prior to each election is highlighted through the black bars. Figure C.17 predicts the sentiment of reports on newspapers, calculated as described in the main part of the paper and as proposed by Proksch et al. (2019). The plot shows the interaction between newspapers and whether a paper endorsed the government.

Figure C.16: Distribution of newspaper partisanship per election in the United Kingdom



Note: The classification is based on information in Wring and Deacon (2010, 2018).

Figure C.17: Predicting the sentiment of coverage on promises in the UK conditional on government endorsement for each newspaper



Note: A linear multilevel regression are used to predict the sentiment. The model reproduces Model 1 of Table 4.4, but adds dummies for each newspaper. The two Sunday newspapers are removed due to low variation in terms of government endorsement (Mail on Sunday) and opposition endorsement (Sunday Mirror). Results do not change when including the Sunday editions. Horizontal bars show 95 per cent confidence intervals.

Appendix D

The Electoral Cycle Effect in Parliamentary Democracies (Supporting Information)

D.1 Descriptive Evidence

Table D.1 shows the number of observations, mean, standard deviation, minimum, median and maximum value for all key dependent, independent and control variables (as well as alternative measurements). Each observations refers to one poll result for a government party. If more than one poll was published for the same party on the same day, the poll results are averaged. We limit our sample to countries where we have at least an average of 10 unique polls per cycle. The mean number of polls per cycle ranges from 13.5 (Belgium) to a maximum of 260 (United Kingdom). Overall, the average number of polls per cycle amounts to 92. Table D.2 lists the availability of polls for all countries included in the analysis.[1] Figure D.1 plots a loess regression line based on all poll results for government parties that are considered

[1]Note that the reported results remain the same when we only consider the 15 countries for which we have at least three cycles (Section D.5).

in the analysis. This loess smoother mirrors closely the shape of the electoral cycle effect in the base model (Figure 5.2a from the paper), indicating that the models are very similar to the aggregated descriptive evidence.

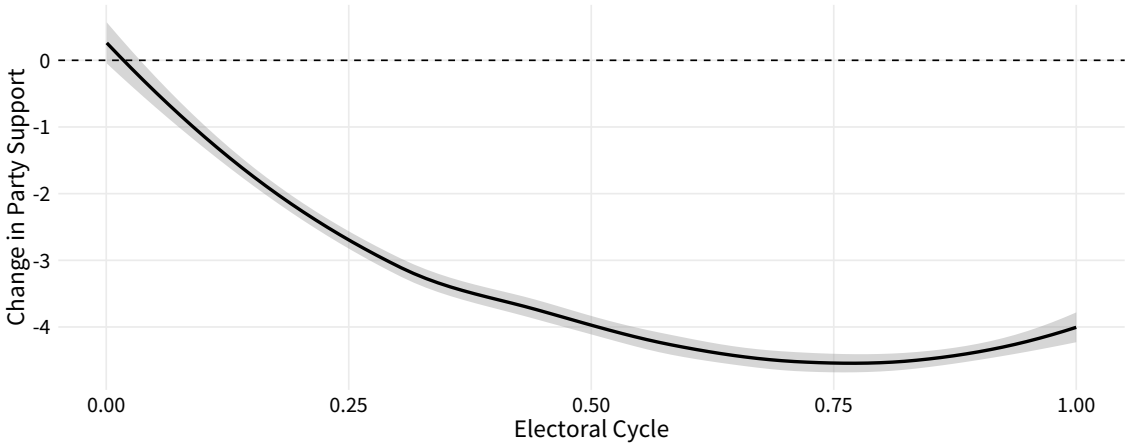
Table D.1: Descriptive statistics

Variable	N	Mean	SD	Min	Median	Max
Poll change	25239	-3.52	6.18	-34.42	-2.69	23.70
Poll change$_{t-1}$	25095	-3.51	6.16	-34.42	-2.65	23.70
Electoral cycle (planned)	25333	0.52	0.28	0.00	0.52	1.00
Electoral cycle (cabinet)	25333	0.55	0.30	0.00	0.57	1.00
Single party government	25333	1.31	0.46	1.00	1.00	2.00
Largest government party	25333	1.68	0.47	1.00	2.00	2.00
Minority government	22884	1.16	0.37	1.00	1.00	2.00
GDP change	25023	0.53	1.15	-7.61	0.59	21.58
GDP change (standardized)	25023	0.56	0.25	0.00	0.60	1.00
Election: Regular	25333	1.70	0.46	1.00	2.00	2.00
Election: Failure	25333	1.13	0.33	1.00	1.00	2.00
Election: Opportunistic	25333	1.18	0.38	1.00	1.00	2.00
PM dissolution power	25333	3.55	4.28	0.00	2.50	10.00
Government dissolution power	25333	1.76	3.40	0.00	0.00	8.50
Party support at last election	25239	31.18	12.84	3.40	35.40	53.66

Table D.2: Cases included in the analysis

Country	Electoral Cycles	Mean Polls Per Cycle	Date of First Poll	Date of Last Poll
Australia	21	62.8	09-12-1960	12-04-2013
Austria	3	90.3	15-07-2006	23-09-2013
Belgium	2	13.5	15-02-2005	07-06-2010
Canada	18	77.3	16-07-1961	18-10-2015
Denmark	16	26.0	15-07-1960	17-06-2015
Finland	1	14.0	15-02-2010	15-03-2011
Germany	15	205.8	15-07-1960	18-09-2013
Greece	4	66.5	15-01-2007	17-09-2015
Hungary	1	19.0	26-11-2009	29-03-2010
Iceland	1	51.0	31-01-2012	26-04-2013
Ireland	12	44.3	28-02-1975	19-02-2016
Italy	1	201.0	17-01-2012	08-02-2013
Japan	5	36.0	09-05-1998	12-12-2012
Netherlands	12	117.0	01-07-1964	11-09-2012
New Zealand	8	69.0	27-11-1987	24-11-2011
Norway	12	91.4	15-10-1965	09-09-2013
Poland	1	94.0	14-08-2010	09-10-2011
Portugal	8	51.2	15-07-1986	02-06-2011
Spain	9	112.7	18-10-1980	20-12-2015
Sweden	3	206.3	13-01-2000	16-09-2010
Turkey	1	27.0	15-06-2010	15-05-2011
United Kingdom	16	260.9	16-08-1955	06-06-2017

Figure D.1: Loess regression line with 95 per cent confidence intervals



D.2 Regression Tables For the Main Models

Models 1–4 from Tables D.3 and D.4 are used to estimate the predicted values of the interaction effects reported in the paper. Figure 5.2a is created based on Model 1, Model 2 predicts the impact of single-party governments in Figure 5.2b, and Model 3 shows the influence of the Prime Minister dissolution power in Figure 5.2c. Model 4 adds an interaction of the respective decade of each observation for the seven parliamentary democracies where polling data are available since the 1960s. The predicted values for the electoral cycle effect in each decade from the 1960s to the 2010s are plotted in Figure 5.3 in the paper.

Table D.3: Multilevel linear regression models of change in party support (main models)

	Model 1	Model 2	Model 3
(Intercept)	1.59 (1.10)	1.84 (1.15)	1.25 (1.07)
El. cycle	-146.92 (3.34)***	-123.79 (3.98)***	-104.63 (4.34)***
El. cycle^2	108.30 (3.24)***	58.72 (3.88)***	58.54 (4.18)***
El. cycle^3	-12.55 (3.21)***	-23.70 (3.81)***	-23.94 (4.11)***
GDP Change	0.01 (0.02)	0.03 (0.02)	0.04 (0.02)
Party support at last election	-0.13 (0.03)***	-0.16 (0.04)***	-0.16 (0.03)***
Election year - 1986	-0.07 (0.02)**	-0.06 (0.02)**	-0.06 (0.02)**
Single party gov.		1.56 (0.96)	
El. cycle × Single party gov.		-104.23 (7.19)***	
El. cycle^2 × Single party gov.		169.79 (6.95)***	
El. cycle^3 × Single party gov.		15.50 (6.88)*	
PM diss. power			0.31 (0.11)**
El. cycle × PM diss. power			-12.95 (0.78)***
El. cycle^2 × PM diss. power			14.77 (0.75)***
El. cycle^3 × PM diss. power			2.35 (0.75)**
Log Likelihood	-65745.91	-65344.40	-65411.81
N	25328	25328	25328
N (Party-Elections)	232	232	232
N (Parties)	62	62	62
N (Countries)	22	22	22

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.
Standard errors in parentheses.

Table D.4: Multilevel linear regression models of change in party support (conditional on decade)

	Model 4
(Intercept)	1.61 (1.45)
El. cycle	−37.34 (16.88)*
El. cycle^2	133.18 (16.61)***
El. cycle^3	65.67 (16.29)***
Decade – 1970s	−0.38 (1.15)
Decade – 1980s	−1.08 (1.17)
Decade – 1990s	−2.80 (1.28)*
Decade – 2000s	−0.79 (1.24)
Decade – 2010s	−1.28 (1.42)
GDP Change	0.07 (0.02)**
Party support at last election	−0.08 (0.03)*
El. cycle × Decade – 1970s	−32.73 (20.24)
El. cycle^2 × Decade – 1970s	−43.48 (19.84)*
El. cycle^3 × Decade – 1970s	−75.52 (19.38)***
El. cycle × Decade – 1980s	−13.72 (18.96)
El. cycle^2 × Decade – 1980s	46.13 (18.69)*
El. cycle^3 × Decade – 1980s	−15.99 (18.36)
El. cycle × Decade – 1990s	−171.52 (18.68)***
El. cycle^2 × Decade – 1990s	36.88 (18.37)*
El. cycle^3 × Decade – 1990s	−57.79 (18.05)**
El. cycle × Decade – 2000s	−60.29 (17.89)***
El. cycle^2 × Decade – 2000s	−108.68 (17.55)***
El. cycle^3 × Decade – 2000s	−93.27 (17.23)***
El. cycle × Decade – 2010s	−94.10 (17.83)***
El. cycle^2 × Decade – 2010s	−16.77 (17.57)
El. cycle^3 × Decade – 2010s	−106.98 (17.26)***
Log Likelihood	-49977.15
N	19624
N (Party-Elections)	151
N (Parties)	28
N (Countries)	7

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.
Standard errors in parentheses.

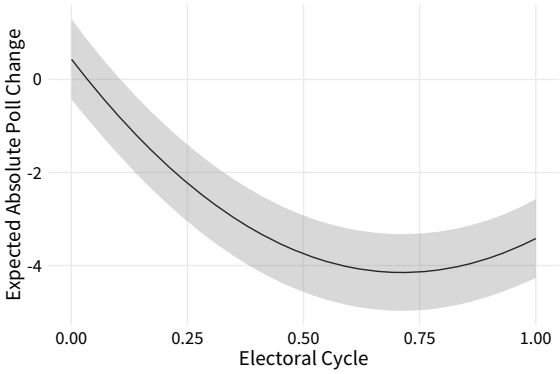
D.3 Regression Assumptions: Autocorrelation and Homoscedasticity

We study party support in opinion polls over an electoral cycle. As a result, the assumption that the residuals in our regression model are independent is potentially problematic due to (serial) autocorrelation. To assess the extent to which this affects our results, we replicate our main models and add an continuous autoregressive correlation structure (CAR1). We opt for a continuous-time specification because our time variable (Electoral cycle) is not integer valued and because the observations are not equally spaced over time. These models are estimated using the `nlme` R package. As it is difficult to specify non-nested structure for the random intercepts in `nlme`, we specify the three grouping levels in our models as nested, but this does not substantially affect our findings. Specifying a CAR(1) correlation structure complicates our model significantly in terms of computational time and feasibility. Particularly when we add interactions to the model, convergence becomes an issue. Therefore we have dropped the cubic time variable ($El.Cycle^3$), which was not significant in the basic model (5), from Models 6–8. We do include the cubic time variable in Model 9 because we expect the electoral cycle effect to diminish in recent decades. Such a decline is captured by the cubic term.

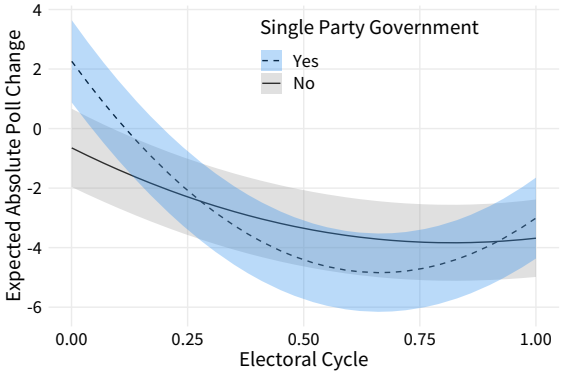
Substantively, the results of our CAR(1) models are comparable to the main models presented in the article even when omitting the cubic term of *Electoral cycle*. The shape of the effect plots is virtually the same for every model (compare Figures D.2 and D.3 with Figures 5.2 and 5.3 printed in the paper).

Figure D.2: Moderated electoral cycle effects (multilevel mixed-effects linear regression models which considers a continuous autoregressive correlation structure (CAR1))

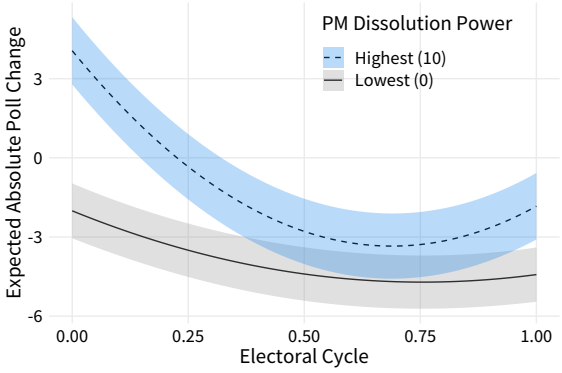
(a) Basic model



(b) Single party government

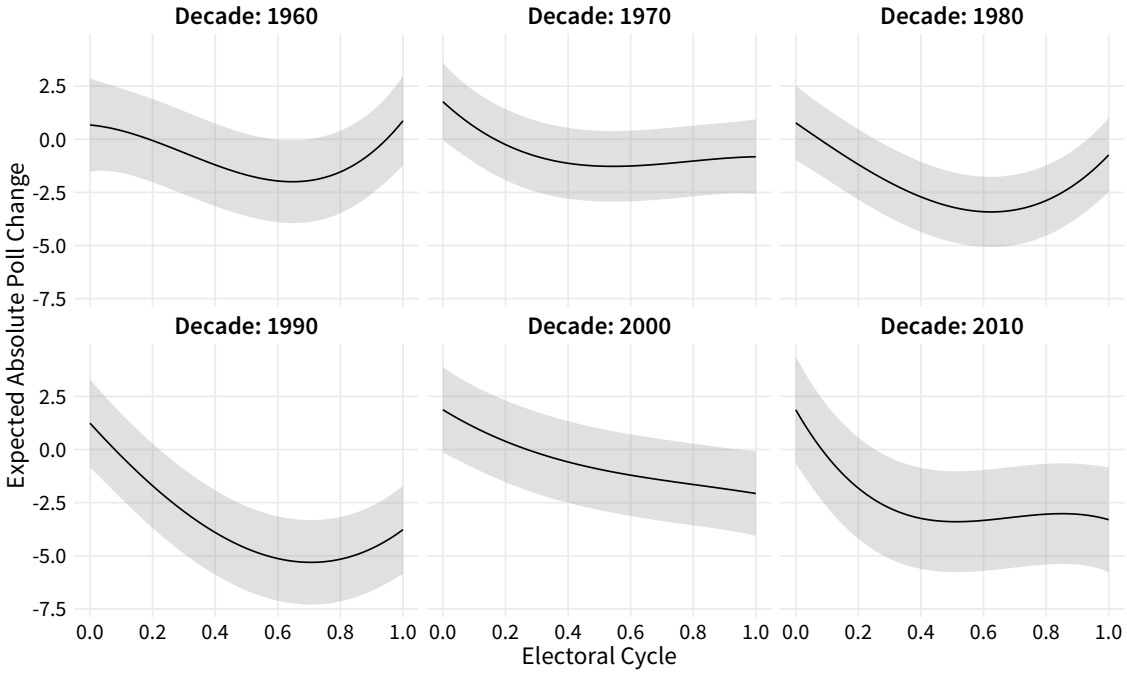


(c) Prime Minister dissolution power



Note: The bands around the lines show the 95 per cent confidence intervals. Figure D.2a is based on Model 5, Figure D.2b on Model 6 of Table D.5, and Figure D.2c on Model 7 of Table D.6.

Figure D.3: Historical development of the electoral cycle effect (multilevel mixed-effects linear regression models which consider the autoregressive correlation structure)



Note: The model only includes countries where we have polling data since the 1960s (Australia, Canada, Denmark, Germany, Netherlands, Norway, United Kingdom). Each panel includes all cycles that *ended* in the respective decade. The gray bands around the lines show the 95 per cent confidence intervals. The Figure is based on Model 9 (Table D.7).

Table D.5: Multilevel mixed-effects linear regression models of change in party support (base model, considering autoregressive correlation structure)

	Model 5	Model 6
El. cycle	-153.951^{***}	
	(5.968)	
El. cycle^2	110.175^{***}	
	(5.753)	
El. cycle^3	-0.671	
	(5.609)	
El. cycle		-153.898^{***}
		(5.951)
El. cycle^2		110.276^{***}
		(5.686)
GDP change	0.025	0.025
	(0.028)	(0.028)
Party support at last election	-0.072^{**}	-0.069^{**}
	(0.025)	(0.025)
Election year - 1986	-0.070^{**}	-0.072^{***}
	(0.022)	(0.021)
Single party gov.	0.146	0.112
	(0.912)	(0.870)
Observations	25,328	25,328
Log Likelihood	$-60{,}218.140$	$-60{,}220.810$
Note:	*p<0.05; **p<0.01; ***p<0.001	

Coefficients with standard errors in parentheses. The regressions replicate Models 1–3, but take into consideration the autoregressive correlation structure (using the nlme R package). Groups (in all models): 22 countries; 171 elections; 232 parties.

Table D.6: Multilevel mixed-effects linear regression models of change in party support (considering autoregressive correlation structure)

	Model 7	Model 8
El. cycle	−129.430***	−99.929***
	(9.735)	(7.949)
El. cycle^2	58.030***	58.009***
	(9.242)	(7.488)
GDP change	0.038	0.037
	(0.034)	(0.027)
Party support at last election	−0.010	−0.106***
	(0.005)	(0.027)
Election year - 1986	−0.067*	−0.060**
	(0.027)	(0.021)
Single party gov.	−0.132	
	(0.861)	
El. cycle × Single party gov.	−70.313***	
	(16.069)	
El. cycle^2 × Single party gov.	139.662***	
	(15.334)	
PM diss. power		0.242**
		(0.083)
El. cycle × PM diss. power		−13.099***
		(1.268)
El. cycle^2 × PM diss. power		13.280***
		(1.217)
(Intercept)	−1.767**	0.073
	(0.656)	(0.854)
Observations	25,328	25,328
Log Likelihood	−61,119.890	−60,088.460
Note:	*p<0.05; **p<0.01; ***p<0.001	

Coefficients with standard errors in parentheses. The regressions replicate Models 1–3, but take into consideration the autoregressive correlation structure (using the nlme R package). Groups (in all models): 22 countries; 171 elections; 232 parties.

Table D.7: Multilevel mixed-effects linear regression models of change in party support (model for decades, considering the autoregressive correlation structure)

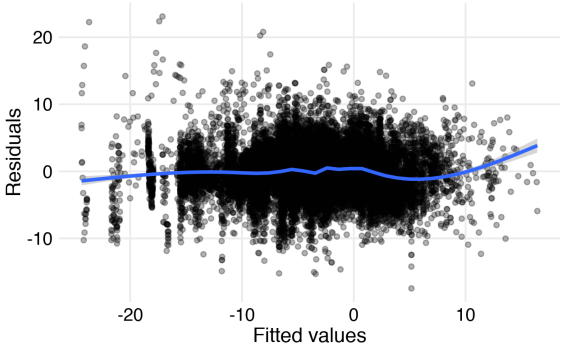
	Model 9
El. cycle	−36.443 (19.802)
El. cycle^2	113.910*** (19.545)
El cycle^3	45.021* (19.056)
Decade – 1970s	0.105 (1.290)
Decade – 1980s	−1.289 (1.277)
Decade – 1990s	−2.818* (1.391)
Decade – 2000s	0.051 (1.378)
Decade – 2010s	−1.769 (1.547)
GDP change	0.061 (0.036)
Party support at last election	−0.049 (0.026)
El. cycle × Decade – 1970s.	−25.622 (24.448)
El. cycle^2 × Decade – 1970s.	−43.601 (24.064)
El. cycle^3 × Decade – 1970s.	−67.089** (23.446)
El. cycle × Decade – 1980s.	−32.510 (23.188)
El. cycle^2 × Decade – 1980s.	27.980 (22.961)
El. cycle^3 × Decade – 1980s.	−22.610 (22.392)
El. cycle × Decade – 1990s.	−162.512*** (23.632)
El. cycle^2 × Decade – 1990s.	33.909 (23.325)
El. cycle^3 × Decade – 1990s.	−32.395 (22.766)
El. cycle × Decade – 2000s	−107.066*** (23.478)
El. cycle^2 × Decade – 2000s	−80.072*** (22.908)
El. cycle^3 × Decade – 2000s	−54.468* (22.427)
El. cycle × Decade – 2010s	−80.719** (25.459)
El. cycle^2 × Decade – 2010s	−9.092 (24.999)
El. cycle^3 × Decade – 2010s	−98.188*** (24.277)
(Intercept)	0.642 (1.365)
Observations	19,624
Log Likelihood	−45,448.730
Note:	*p<0.05; **p<0.01; ***p<0.001

Coefficients with standard errors in parentheses. The regression replicates Models 4, but takes into consideration the autoregressive correlation structure (using the nlme R package). Groups: 7 countries; 109 elections; 151 parties.

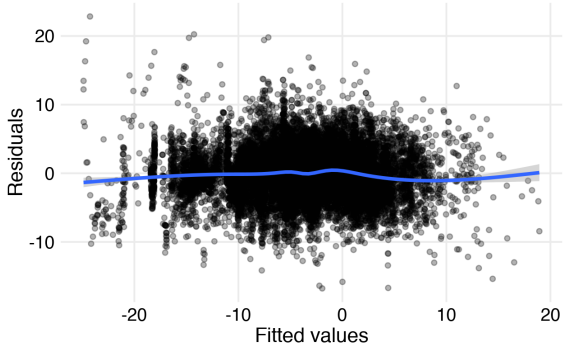
To test whether the standard deviations of the error terms are constant and independent across all values, we plot the predicted values against the fitted values (Figure D.4). We use the base model (Model 1), the combined model that includes the relevant interactions (Model A5) and the base model that adds $Poll\ Change_{t-1}$ (Model A8). To each scatterplot a GAM smoother is added. The plots show that the models do not seem to be biased due to heteroscedasticity.

Figure D.4: Plotting residuals against fitted values for Model 1, Model A5 and Model A8

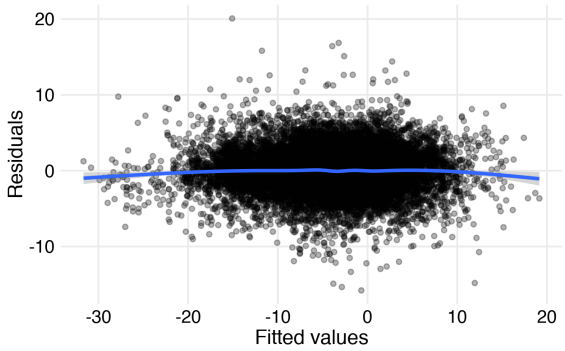
(a) Basic model (M1)



(b) Combined Model (M A5)



(c) Basic Model with lagged DV (M A8)



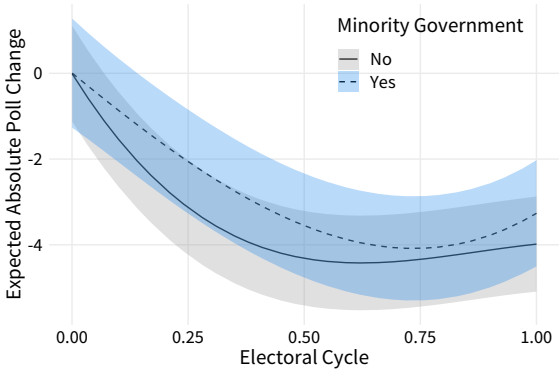
D.4 Additional Institutional Features

The paper shows plots with predicted values for three key moderators of the electoral cycle effect: (1) single-party government vs coalition government, (2) prime minister dissolution power, and (3) the development over time. In this section we show how additional institutional and economic variables influence the shape of the electoral cycle effect. The regression tables that were used to estimate the predicted values are printed below, and the captions of the plots indicate which model has been used for the prediction.

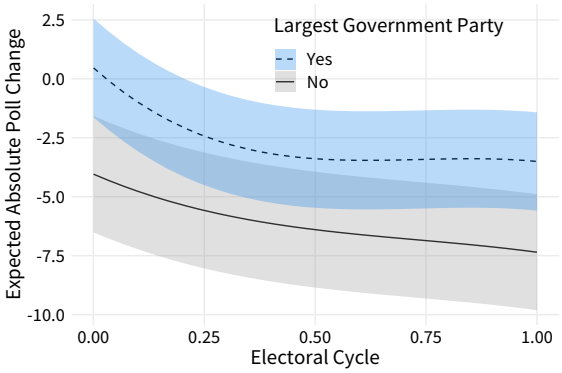
Figure D.5a, based on Model A1, checks whether the electoral cycle effect differs for minority governments compared to majority governments, but the shape of the effects are extremely similar. Model A2 tests whether being the largest party in a coalition affects the electoral cycle effect. Again, the results are inconclusive (Figure D.5b). Model A3 uses Lijphart's (2012) executive-parties and federal-unitary dimensions as proxies for clarity of responsibility. As expected, the electoral cycle effect is more pronounced in countries with a strong executive and in unitary democracies (Figures D.5c and D.5d). Model A4 uses the dissolution power of the government (instead of the dissolution power of the prime ministerial party which is reported in the main part of the paper). The scores for government dissolution power (theoretically ranging from 0 to 10) are also derived from Goplerud and Schleiter (2016). The shape of the cycles is very similar when we set the dissolution powers to the lowest and highest value (Figure D.5e). While we observe an effect for strong prime ministerial control over the electoral cycle (Figure 5.2c), we do not see a similar effect for countries with strong government dissolution powers.

Figure D.5: Moderated electoral cycle effects

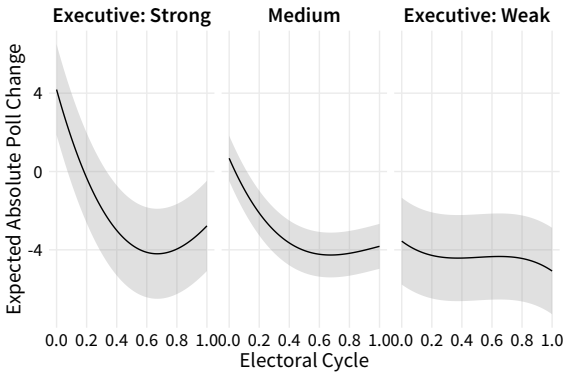
(a) Minority government (M A1)



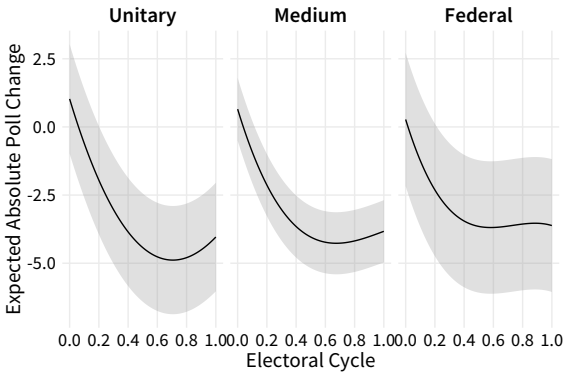
(b) Largest government party (M A2)



(c) Executive-Parties dimension (M A3)



(d) Federal-Unitary dimension (M A3)



(e) Government dissolution power (M A4)

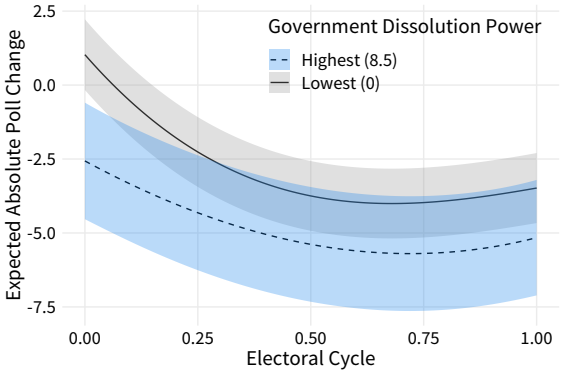


Table D.8: Multilevel linear regression models of change in party support (conditional on minority government and largest government party)

	Model A1	Model A2
(Intercept)	1.31	−0.15
	(1.18)	(1.36)
El. cycle	−123.98***	−110.80***
	(4.04)	(4.24)
El. cycle^2	106.85***	26.35***
	(3.80)	(4.14)
El. cycle^3	−25.80***	−10.36*
	(3.77)	(4.10)
GDP Change	0.10***	−0.06**
	(0.02)	(0.02)
Party support at last election	−0.13***	−0.17***
	(0.03)	(0.05)
Election year - 1986	−0.07**	−0.09***
	(0.02)	(0.02)
Minority gov.	0.61	
	(0.36)	
El. cycle × Minority gov.	−25.15*	
	(10.25)	
El. cycle^2 × Minority gov.	−9.50	
	(9.53)	
El. cycle^3 × Minority gov.	38.56***	
	(9.34)	
Largest gov. party		3.42**
		(1.20)
El. cycle × Largest gov. party		7.75
		(5.75)
El. cycle^2 × Largest gov. party		44.23***
		(5.60)
El. cycle^3 × Largest gov. party		−17.09**
		(5.53)
Log Likelihood	-59431.04	-42723.80
N	22767	17530
N (Party-Elections)	229	142
N (Parties)	61	48
N (Countries)	22	17

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.
Standard errors in parentheses.

Table D.9: Multilevel linear regression models of change in party support (conditional on Lijphart index and government dissolution power)

	Model A3	Model A4
(Intercept)	2.55 (1.47)	2.51 (1.23)*
El. cycle	−171.07 (3.37)***	−157.49 (3.73)***
El. cycle^2	129.73 (3.28)***	118.64 (3.64)***
El. cycle^3	−17.76 (3.26)***	−17.96 (3.62)***
GDP Change	0.04 (0.02)*	0.01 (0.02)
Party support at last election	−0.16 (0.04)***	−0.15 (0.03)***
Election year - 1986	−0.07 (0.02)**	−0.07 (0.02)**
Executive-parties dim.	−0.72 (0.74)	
Federal-unitary dim.	0.13 (0.48)	
El. cycle × Executive-parties dim.	79.58 (3.79)***	
El. cycle^2 × Executive-parties dim.	−80.95 (3.62)***	
El. cycle^3 × Executive-parties dim.	−2.58 (3.59)	
El. cycle × Federal-unitary dim.	20.51 (2.48)***	
El. cycle^2 × Federal-unitary dim.	−13.12 (2.41)***	
El. cycle^3 × Federal-unitary dim.	−7.00 (2.39)**	
Gov. diss. power		−0.23 (0.14)
El. cycle × Gov. diss. power		5.86 (1.01)***
El. cycle^2 × Gov. diss. power		−5.11 (0.95)***
El. cycle^3 × Gov. diss. power		2.42 (0.93)**
Log Likelihood	-64305.52	-65702.82
N	25090	25328
N (Party-Elections)	228	232
N (Parties)	58	62
N (Countries)	19	22

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.
Standard errors in parentheses.

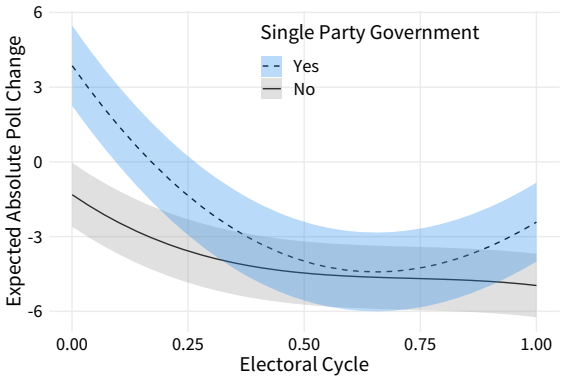
D.5 Combined Models and Focussing on Selected Countries

In the main part of the paper, we do not add all independent variables to one model to avoid multicollinearity. However, as we show in this section, the shapes of the estimated electoral cycle effects remain the same when we add both *Single party government* and *Prime Minister dissolution power* to the basic model (Model 1). Figure D.6 shows the plots from the paper and contrasts them to the plots estimated from the combined model (Model A5).

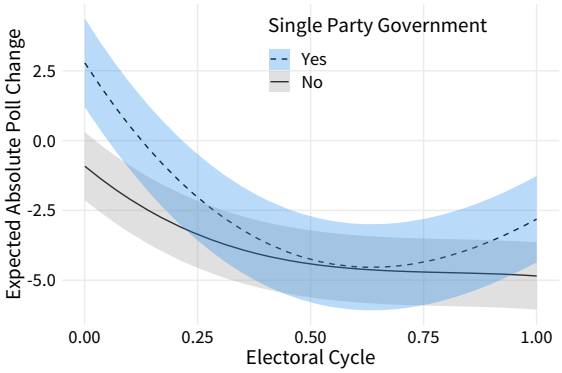
Figure D.7 contrasts the results from the paper with a combined model that is restricted only to the countries with data for three or more electoral cycles (Model A6). Again, the shapes of the effects remain very similar.

Figure D.6: Comparing the main effects with a combined model

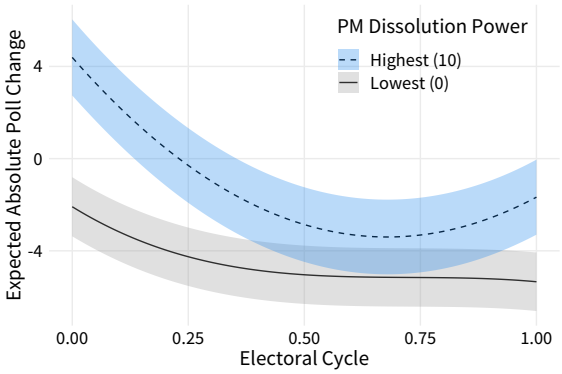
(a) Single party government (M 2)



(b) Single party government (combined; M A5)



(c) Prime Minister dissolution power (M 3)



(d) Prime Minister dissolution power (combined; M A5)

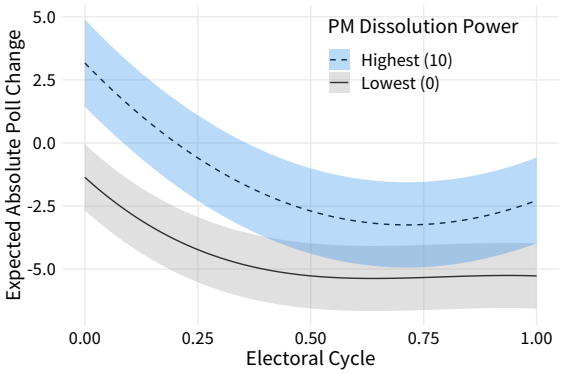
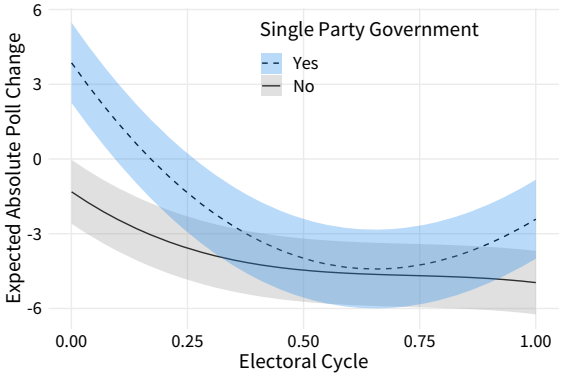
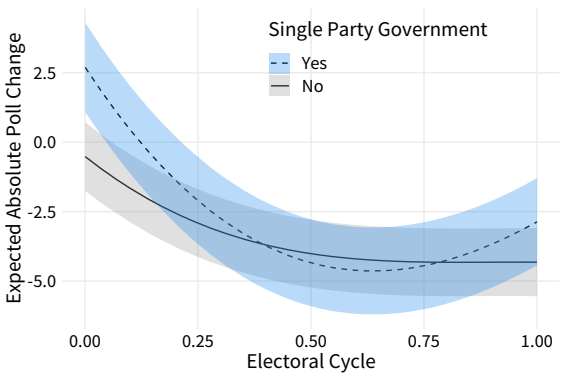


Figure D.7: Comparing the main effects with a combined model that only considers countries with at least three electoral cycles

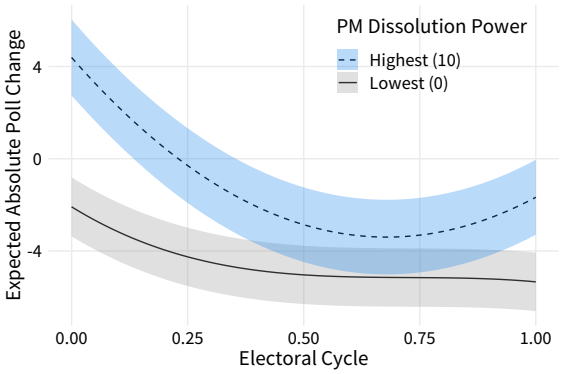
(a) Single party government (M 2)



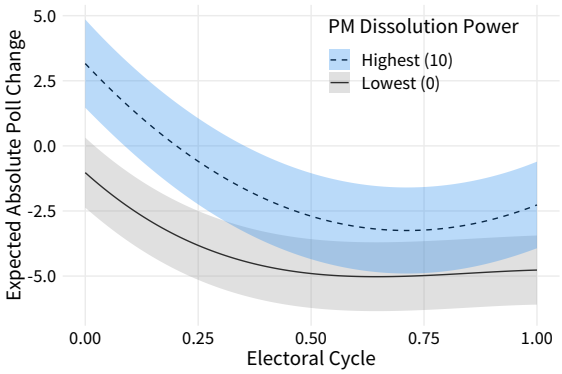
(b) Single party government (subset); M A6



(c) Prime Minister dissolution power (M 3)



(d) Prime Minister dissolution power (subset); M A6

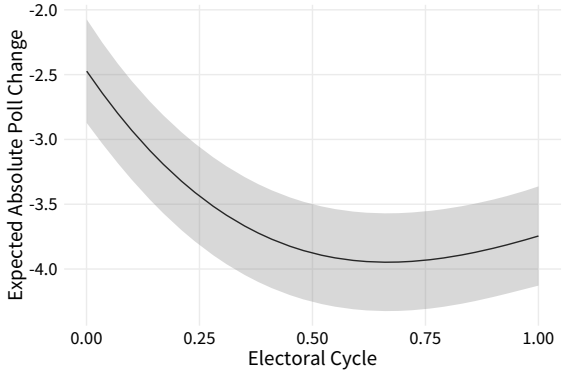


D.6 Alternative Operationalizations of Poll Change

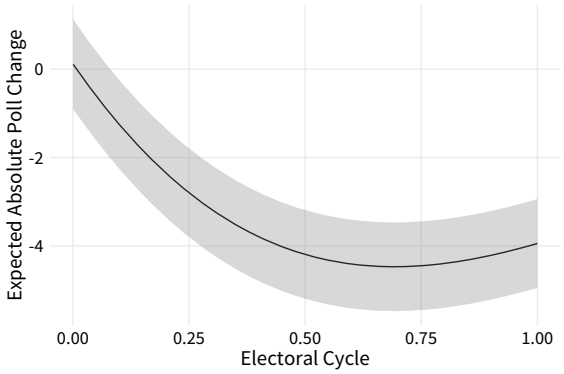
In the main part of the paper we use the *Absolute poll change* as the dependent variable. Here we show that the substantive conclusions do not change when we add the lagged *Poll change* (Figure D.8a based on Model A7) or use the *Percentage poll change* as the dependent variable (Figure D.8c based on Model A9). We decided to use the *Absolute poll change* as the main dependent variable because the absolute poll changes most closely mirror the descriptive evidence from the *loess* regression (Figure D.1).

Figure D.8: Comparing the electoral cycle effect with and without the lagged dependent variable (DV) (i.e., absolute poll change), and with the percentage poll change

(a) Basic model with lagged DV (M A8)



(b) Basic model without lagged DV (M A7)



(c) Percentage poll change as DV (M A9)

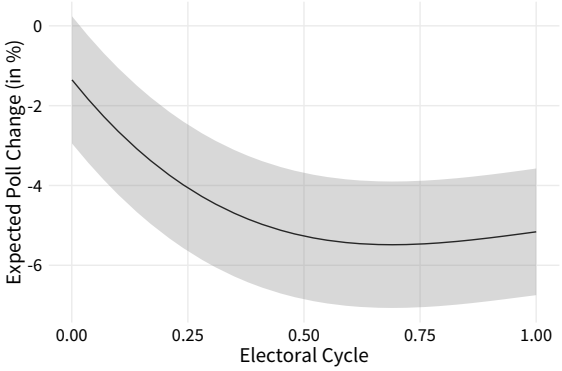


Table D.10: Multilevel linear regression models of change in party support (based on subset of countries)

	Model A5	Model A6
(Intercept)	1.50	1.23
	(1.09)	(1.09)
El. cycle	-106.40^{***}	-103.52^{***}
	(4.39)	(4.36)
El. cycle^2	47.15^{***}	52.14^{***}
	(4.25)	(4.28)
El. cycle^3	-29.33^{***}	-24.02^{***}
	(4.17)	(4.21)
GDP Change	0.04^{*}	0.04^{*}
	(0.02)	(0.02)
Party support at last election	-0.18^{***}	-0.16^{***}
	(0.04)	(0.04)
Election year - 1986	-0.06^{*}	-0.05^{*}
	(0.02)	(0.02)
PM diss. power	0.29^{**}	0.25^{*}
	(0.11)	(0.11)
Single party gov.	1.06	0.55
	(0.95)	(0.96)
El. cycle × PM diss. power	-9.31^{***}	-9.54^{***}
	(0.95)	(0.95)
El. cycle^2 × PM diss. power	6.47^{***}	6.09^{***}
	(0.92)	(0.92)
El. cycle^3 × PM diss. power	2.51^{**}	2.09^{*}
	(0.91)	(0.92)
El. cycle × Single party gov.	-51.29^{***}	-54.65^{***}
	(8.87)	(8.82)
El. cycle^2 × Single party gov.	133.38^{***}	130.47^{***}
	(8.49)	(8.56)
El. cycle^3 × Single party gov.	4.37	1.95
	(8.35)	(8.51)
Log Likelihood	-65266.35	-63782.26
N	25328	24763
N (Party-Elections)	232	221
N (Parties)	62	52
N (Countries)	22	15

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.
Standard errors in parentheses. Model 6 only includes countries with data for at least three cycles.

Table D.11: Multilevel linear regression models of change in party support (testing for lagged dependent variable)

	Model A7	Model A8	Model A9
(Intercept)	1.59 (1.10)	0.62 (0.41)	−0.64 (1.50)
El. cycle	−146.92 (3.34)***	−43.55 (2.59)***	−133.69 (3.56)***
El. cycle^2	108.30 (3.24)***	36.00 (2.47)***	94.93 (3.45)***
El. cycle^3	−12.55 (3.21)***	−4.65 (2.40)	−16.63 (3.41)***
GDP Change	0.01 (0.02)	0.01 (0.01)	−0.00 (0.02)
Party support at last election	−0.13 (0.03)***	−0.05 (0.01)***	−0.10 (0.04)*
Election year - 1986	−0.07 (0.02)**	−0.02 (0.01)**	−0.07 (0.02)***
Poll change$_{t-1}$		0.67 (0.00)***	
Log Likelihood	-65745.91	-57922.56	-67132.56
N	25328	25203	25328
N (Party-Elections)	232	230	232
N (Parties)	62	62	62
N (Countries)	22	22	22

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.
Standard errors in parentheses. Model A7 replicates Model 1, Model A8 adds the lagged DV, Model A9 uses the percentage change of a polling result compared to the previous election.

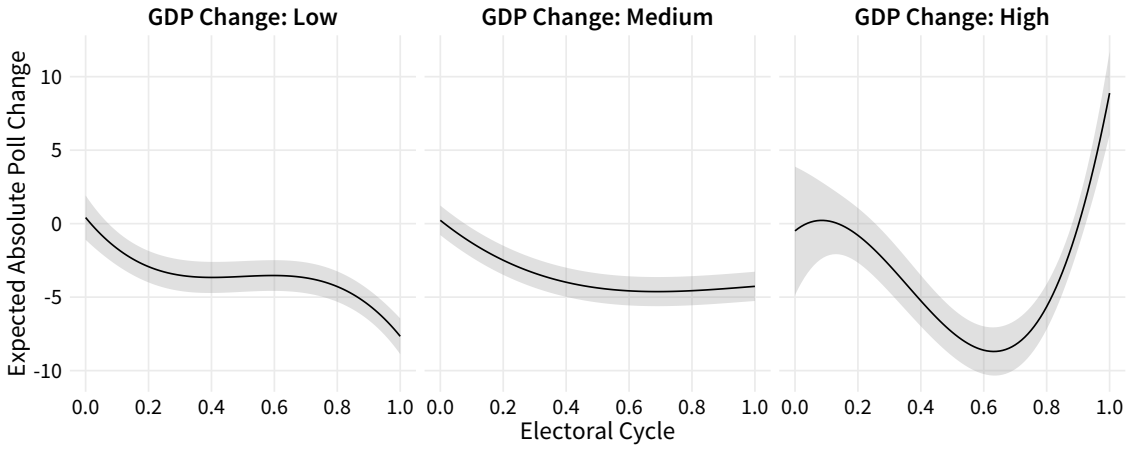
D.7 Economic Performance and the Electoral Cycle Effect

Additionally, we how the GDP growth compared to the previous quarter (lagged by one quarter) affects the shape of the electoral cycle. Recall that we control for *GDP change* in all models in the paper. Thus, we analyse the shape of the electoral cycle effect after controlling for economic performance. We also check how low, medium, and high GDP change impacts the electoral cycle effect by adding an interaction *El. cycle*, *El. cycle*^{2} and *El. cycle*^{3} with the quarterly *GDP Change* (Figure D.9). The results based on Table D.12 are as expected from the economic voting literature: if the economy performs badly prior to an election (left-hand panel), government parties are expected to lose support dramatically before the election. However, if the economy performs very well (indicated by a large positive GDP Change), we observe a strong upturn in support for government parties (right-hand panel).

Additionally, we standardise *GDP Change* by country and decade in order to explore whether the effects of economic growth are dependent on the country-specific situation rather than absolute levels of growth. The quarter with the lowest GDP growth in each decade for each country gets the value 0, the highest value gets the value 1. This regression thus includes the relative economic performance to control for differences between countries and over time (Model A11). The curvilinear relationship for periods with high GDP growth becomes much smaller (Figure D.10).

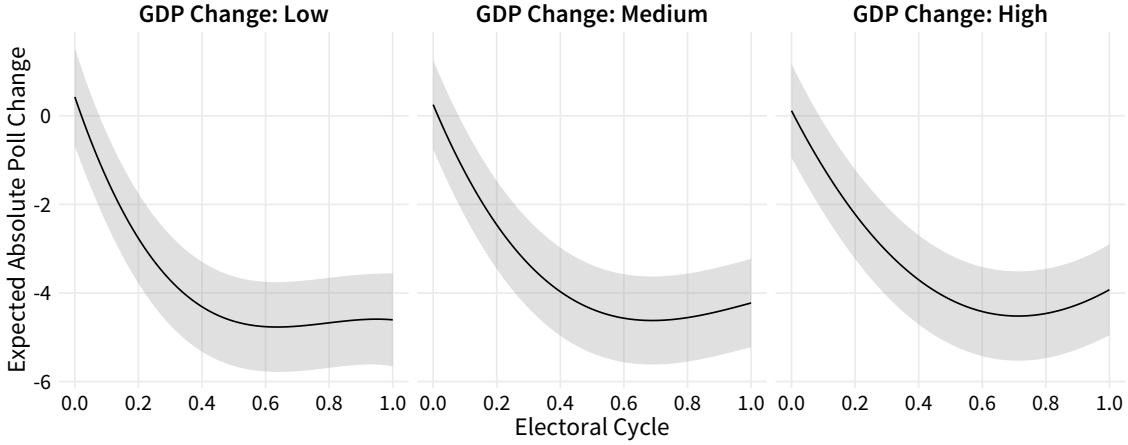
Taking the results from this section together, clearly shows that economic conditions matter for the electoral cycle effect. At the same time, the effects we report in this paper and the appendix exist *even* when controlling for economic performance.

Figure D.9: The moderating impact of quarterly GDP growth and the electoral cycle effect



Note: The model interacts *El. cycle*, *El. cycle*^{2} and *El. cycle*^{3} with the quarterly *GDP Change*. GDP change is the percentage point change of the GDP (lagged by one quarter). Results are based on Model A10.

Figure D.10: The moderating impact of standardised quarterly GDP growth and the electoral cycle effect



Note: The model interacts *El. cycle*, *El. cycle*^{2} and *El. cycle*^{3} with *GDP Change*, the change of the standardised GDP compared to previous quarter (lagged by one quarter). *GDP Change* is standardised for each country and decade where 0 is the lowest and 1 the highest value of GDP change in the respective country-decade subset. Results are based on Model A11.

Table D.12: Multilevel linear regression models of change in party support (testing different measures of GDP change)

	Model A10	Model A11
(Intercept)	1.47 (1.09)	1.32 (1.10)
El. cycle	-113.20 (4.71)***	-108.92 (9.31)***
El. cycle^2	34.96 (4.62)***	51.84 (8.82)***
El. cycle^3	-37.36 (4.53)***	-45.88 (8.79)***
GDP Change	0.04 (0.02)*	-0.03 (0.03)
Party support at last election	-0.18 (0.04)***	-0.18 (0.04)***
Election year - 1986	-0.06 (0.02)**	-0.06 (0.02)**
PM diss. power	0.29 (0.11)**	0.29 (0.11)**
Single party gov.	1.03 (0.95)	1.05 (0.95)
El. cycle × GDP Change	13.27 (3.25)***	
El. cycle^2 × GDP Change	25.86 (3.45)***	
El. cycle^3 × GDP Change	17.00 (3.26)***	
El. cycle × PM diss. power	-9.81 (0.96)***	-9.23 (0.96)***
El. cycle^2 × PM diss. power	5.59 (0.92)***	6.47 (0.92)***
El. cycle^3 × PM diss. power	1.85 (0.91)*	2.48 (0.91)**
El. cycle × Single party gov.	-49.87 (8.87)***	-51.45 (8.89)***
El. cycle^2 × Single party gov.	135.10 (8.49)***	132.48 (8.53)***
El. cycle^3 × Single party gov.	5.51 (8.36)	4.19 (8.37)
GDP Change (standardised)		0.40 (0.15)**
El. cycle × GDP Change (standardised)		1.62 (14.22)
El. cycle^2 × GDP Change (standardised)		-7.05 (13.36)
El. cycle^3 × GDP Change (standardised)		29.96 (13.57)*
Log Likelihood	-65210.74	-65250.34
N	25328	25328
N (Party-Elections)	232	232
N (Parties)	62	62
N (Countries)	22	22

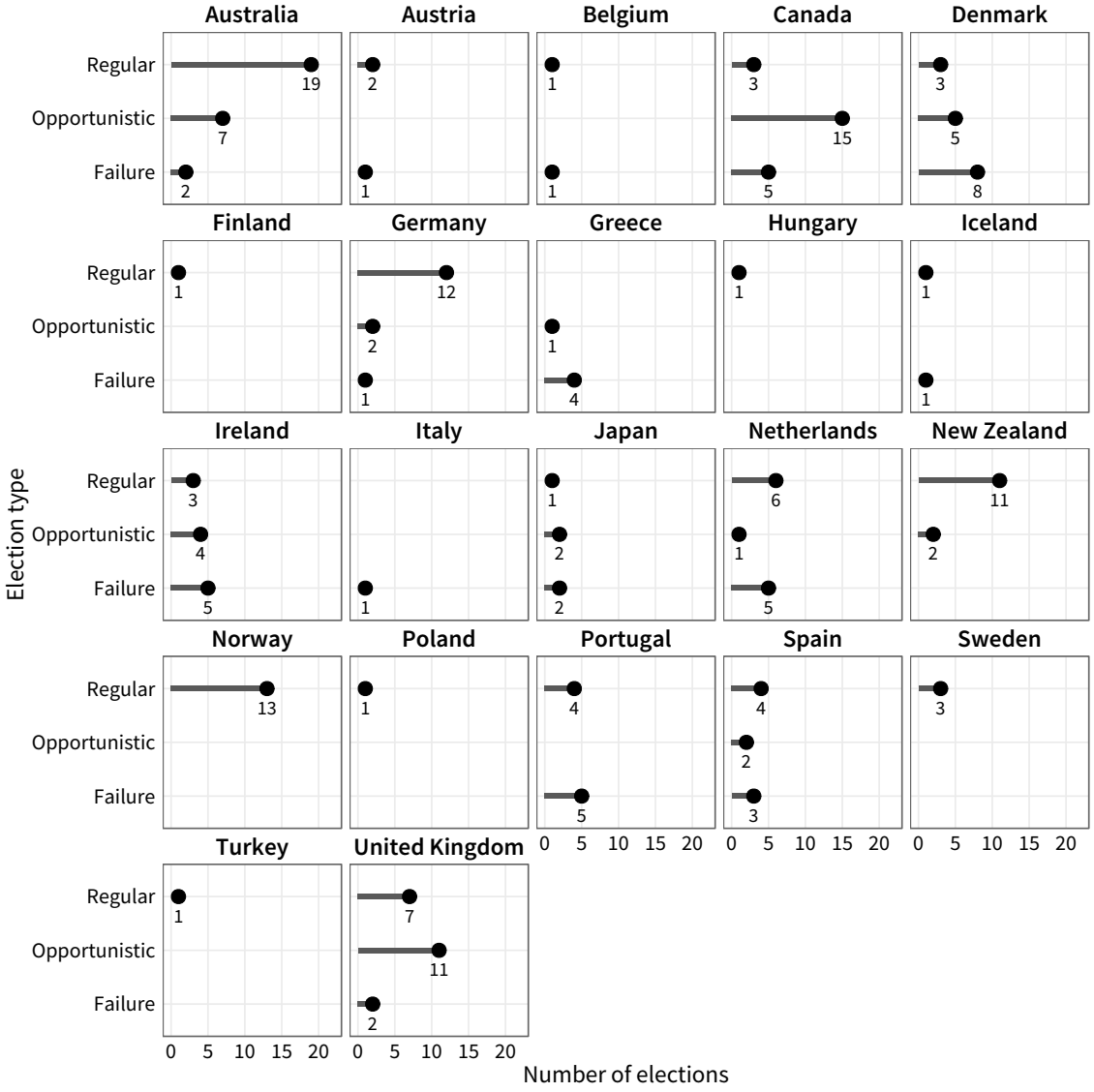
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.
Standard errors in parentheses. Model A10 adds interaction of GDP Change and El. cycle to Model A5, Model A11 uses a measure of GDP change which is standardised by country and decade.

D.8 The End of an Electoral Cycle

In the paper, we analysed all parliamentary elections in our dataset, provided that the time between elections was at least half a year. We have not distinguished between regular and early (opportunistic or failure) elections. This coding is based on the dataset by Schleiter and Tavits (2016) and additional data collected by the authors. When an election lasts the maximum duration and is not affected by cabinet reshuffles, we regard the period as a *Regular election.* If the government takes the initiative to call an early election and opportunistic behaviour by the incumbent government is observable, the cycle falls under the category *Opportunistic election.* A *Failure election* is either forced by the opposition or when the government fails due to, a lost vote of confidence, the resignation of a coalition party, the resignation of the prime minister. We collected additional information on the countries and elections missing from that dataset, based (in order of preference) on election reports in scientific journals, other scientific sources, newspaper reports from NexisLexis and Wikipedia. Figure D.11 shows the number of elections analysed for each country, faceted by the election type (*Regular/Opportunistic/Failure*). Note that the results do not change if we exclude countries with fewer than three cycles (Section D.5).

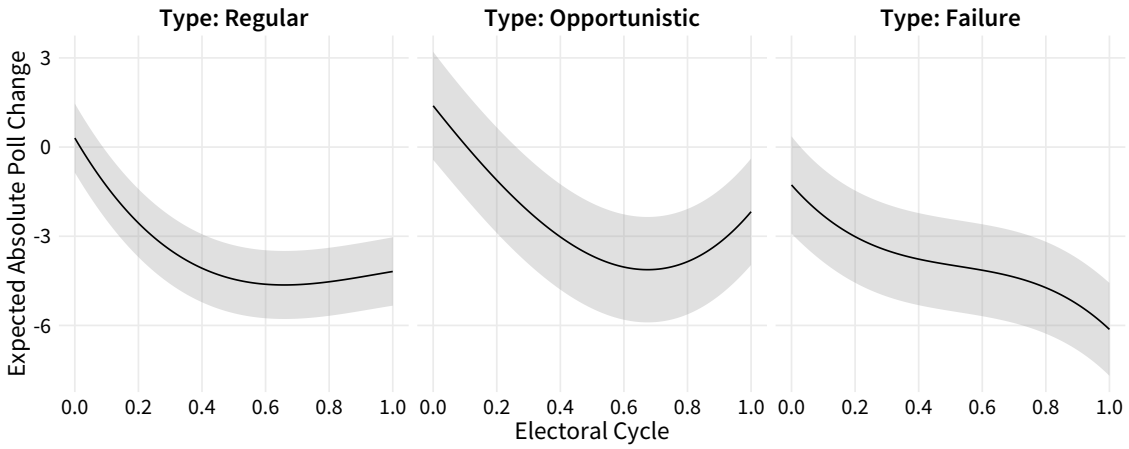
Figure D.12 displays the electoral cycle for three types of elections based on Model A12: regular elections, failure elections (where the government lost parliamentary support and was forced to resign) and opportunistic elections (where the government chose to resign early). Regular elections (left panel of Figure D.12) show a modest u-shaped electoral cycle effect with losses early in the term and a slight recovery (but no rise) from about midway through the cycle. Opportunistic elections (middle panel) show a slightly different pattern: these governments start out with higher levels of support at the very start of the term, but then quickly lose support during the first half of the term. Only at the very end of the term we observe a somewhat of a recovery. Note that in these cases the timing of the elections is controlled by

Figure D.11: The number of elections for each country, faceted by election type



the government (or the PM alone). Perhaps the best governments can do is avoid further losses by calling early elections or they might be too optimistic about their chances of influencing the election result by controlling the timing. There are enough examples of governments calling early elections when polling well, but then losing support during the campaign. Early elections following government failure show an entirely different pattern from the previous two types: these cycles are characterised by a steady, almost linear, decline of government party support.

Figure D.12: The moderating impact of the election type and the electoral cycle effect



Note: The plot shows the interaction of *El. cycle*, *El. cycle*^{2} and *El. cycle*^{3} with the type of election (Regular, Opportunistic, or Failure). For the classification of the elections see Schleiter and Tavits (2016). Results are based on Model A12.

D.9 A Different Measurement of the End of the Electoral Cycle

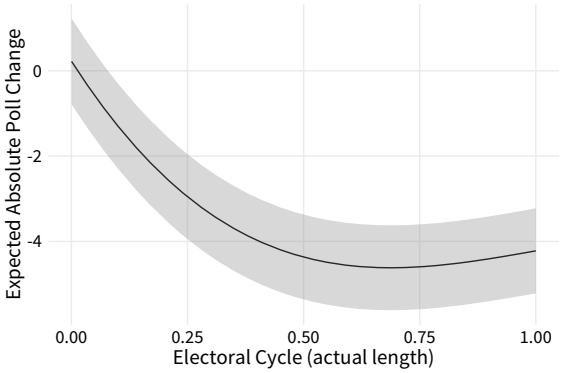
Beside the nominal measure of the end of an electoral cycle, we also check whether the results persists when we use a different specification of *Electoral cycle*. The main measure ranges between 0 (date of cabinet inauguration) to 1 (date of next election). A concern might be that this measure is endogenous to poll support: governments may call early elections when their poll numbers are favourable.

An alternative, exogenous measure of the point in the electoral cycle, uses the 'planned' end date of the cycle. For instance, if elections are scheduled every four years, *Electoral cycle* would take the value of 1 four years after the previous election. If a legislative period was interrupted by an early election after two years, *Electoral cycle* would end at 0.5. Model A14 uses this alternative measurement of *Electoral cycle*. Comparing the coefficients from Model A13 and Model A14 shows that in both models the size, direction and standard errors of the coefficients for *El. cycle*, *El. cycle*^{2} and *El. cycle*^{3} are very similar to the respective coefficients for *El. cycle (planned)*, *El. cycle*^{2} *(planned)* and *El. cycle*^{3} *(planned)*. The coefficients for the other independent variables and interactions across Models A13 and A14 are almost identical. Figure A13 compares the shape of the electoral cycle based on Model A13 (Figure D.13a) and Model A14 (Figure D.13b). The shapes are very similar. Thus, the results are not driven by our main measurement of the point of an observation during the electoral cycle.

We nevertheless prefer our measure based on the actual cycle, because the points in the cycle are better comparable across elections: all cycles start at 0 and end at 1, whereas in a measure based on planned length some cycles will have elections mid-way and end there. Future work could look into this issue in more detail by analysing the moment at which early elections were called in more detail.

Figure D.13: Comparing the electoral cycle effect conditional on different measurements of the end of the cycle

(a) Basic model with actual length (M A13)



(b) Basic model with planned length (M A14)

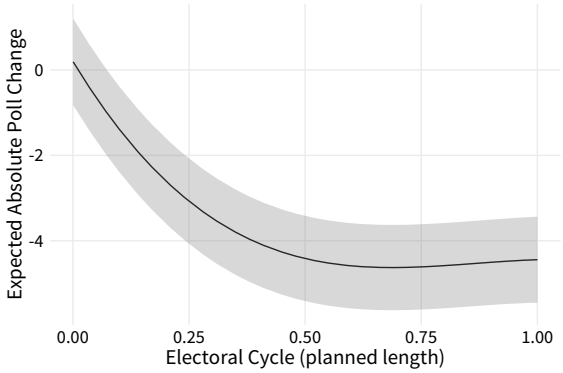


Table D.13: Multilevel linear regression models of change in party support (conditional on type of election and alternative measure of electoral cycle)

	Model A12	Model A13	Model A14
(Intercept)	1.33 (1.23)	1.50 (1.09)	1.12 (1.10)
El. cycle	−127.84 (10.99)***	−106.40 (4.39)***	
El. cycle^2	−61.78 (10.47)***	47.15 (4.25)***	
El. cycle^3	−45.28 (10.13)***	−29.33 (4.17)***	
El. – Opportunistic	1.43 (0.99)		
El. – Regular	0.36 (0.84)		
Single party gov.	1.01 (0.95)	1.06 (0.95)	0.68 (0.94)
GDP Change	0.03 (0.02)	0.04 (0.02)*	0.05 (0.02)*
Party support at last election	−0.19 (0.04)***	−0.18 (0.04)***	−0.17 (0.04)***
Election year - 1986	−0.05 (0.02)*	−0.06 (0.02)*	−0.04 (0.02)*
PM diss. power	0.26 (0.11)*	0.29 (0.11)**	0.25 (0.11)*
El. cycle × El. – Opportunistic	19.42 (12.76)		
El. cycle^2 × El. – Opportunistic	163.31 (12.16)***		
El. cycle^3 × El. – Opportunistic	58.47 (12.05)***		
El. cycle × El. – Regular	24.91 (10.95)*		
El. cycle^2 × El. – Regular	114.54 (10.44)***		
El. cycle^3 × El. – Regular	15.57 (10.17)		
El. cycle × PM diss. power	−9.20 (0.98)***	−9.31 (0.95)***	
El. cycle^2 × PM diss. power	5.81 (0.95)***	6.47 (0.92)***	
El. cycle^3 × PM diss. power	1.53 (0.94)	2.51 (0.91)**	
El. cycle × Single party gov.	−49.14 (8.97)***	−51.29 (8.87)***	
El. cycle^2 × Single party gov.	143.55 (8.59)***	133.38 (8.49)***	
El. cycle^3 × Single party gov.	3.26 (8.43)	4.37 (8.35)	
El. cycle (planned)			−115.53 (4.47)***
El. cycle^2 (planned)			44.03 (4.31)***
El. cycle^3 (planned)			−24.69 (4.24)***
El. cycle (planned) × PM diss. power			−6.60 (1.03)***
El. cycle^2 (planned) × PM diss. power			11.17 (0.94)***
El. cycle^3 (planned) × PM diss. power			1.29 (0.92)
El. cycle (planned) × Single party gov.			−48.21 (9.54)***
El. cycle^2 (planned) × Single party gov.			55.90 (8.55)***
El. cycle^3 (planned) × Single party gov.			−12.00 (8.31)
Log Likelihood	-65129.41	-65266.35	-64382.21
N	25328	25328	24929
N (Party-Elections)	232	232	232
N (Parties)	62	62	62
N (Countries)	22	22	22

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Standard errors in parentheses. Model A12 includes the type of election, Model A13 reproduces Model A5 to compare the coefficients to Model A14 that uses the planned length of the electoral cycle for each country.

Bibliography

Achen, Christopher H., and Larry M. Bartels. 2016. *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton: Princeton University Press.

Adams, James, and Samuel Merrill III. 2009. "Policy-Seeking Parties in a Parliamentary Democracy with Proportional Representation: A Valence-Uncertainty Model." *British Journal of Political Science* 39 (3): 539–558.

Adams, James, and Zeynep Somer-Topcu. 2009. "Moderate Now, Win Votes Later: The Electoral Consequences of Parties' Policy Shifts in 25 Postwar Democracies." *The Journal of Politics* 71 (2): 678–692.

Adams, James, Michael Clark, Lawrence Ezrow, and Garrett Glasgow. 2006. "Are Niche Parties Fundamentally Different from Mainstream Parties? The Causes and the Electoral Consequences of Western European Parties' Policy Shifts, 1976–1998." *American Journal of Political Science* 50 (3): 513–529.

Adcock, Robert, and David Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *American Political Science Review* 95 (3): 529–546.

Alexiadou, Despina, and Danial Hoepfner. 2018. "Platforms, Portfolios, Policy: How Audience Costs Affect Social Welfare Policy in Multiparty Cabinets." *Political Science Research and Methods* published ahead of print (doi: 10.1017/psrm.2018.2).

American Political Science Association. 1950. "Toward a More Responsible Two-Party System: A Report of the Committee on Political Parties." *American Political Science Review* 44 (3): v–96.

Anderson, Christopher J. 1995. "The Dynamics of Public Support for Coalition Governments." *Comparative Political Studies* 28 (3): 350–383.

Andersson, Staffan, Torbjörn Bergman, and Svante Ersson. 2014. *The European Representative Democracy Data Archive, Release 3: Main Sponsor: Riksbankens Jubileumsfond (In2007-0149:1-E).* URL: http://www.erdda.se.

Arnold, R. Douglas. 2004. *Congress, the Press, and Political Accountability.* Princeton: Princeton University Press.

Artés, Joaquín. 2013. "Do Spanish Politicians Keep Their Promises?." *Party Politics* 19 (1): 143–156.

Artés, Joaquín, and Antonio Bustos. 2008. "Electoral Promises and Minority Governments: An Empirical Study." *European Journal of Political Research* 47 (3): 307–333.

Bäck, Hanna, Marc Debus, and Heike Klüver. 2016. "Bicameralism, Intra-Party Bargaining, and the Formation of Party Policy Positions: Evidence from the German Federal System." *Party Politics* 22 (3): 405–417.

Bäck, Hanna, Markus Baumann, Marc Debus, and Jochen Müller. 2019. "The Unequal Distribution of Speaking Time in Parliamentary-Party Groups." *Legislative Studies Quarterly* 44 (1): 163–193.

Ban, Pamela, Alexander Fouirnaies, Andrew B. Hall, and James M. Snyder. 2018. "How Newspapers Reveal Political Power." *Political Science Research and Methods* published ahead of print (doi: 10.1017/psrm.2017.43).

Bates, Douglas, Martin Mächler, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67 (1): 1–48.

Baumeister, Roy F., Ellen Bratslavsky, and Catrin Finkenauer. 2001. "Bad Is Stronger Than Good." *Review of General Psychology* 5 (4): 323–370.

Bélanger, Éric, and Bonnie M. Meguid. 2008. "Issue Salience, Issue Ownership, and Issue-Based Vote Choice." *Electoral Studies* 27 (3): 477–491.

Belchior, Ana Maria. 2018. "The Effects of Party Identification on Perceptions of Pledge Fulfilment: Evidence from Portugal." *International Political Science Review* published ahead of print (doi: 10.1177/0192512118791298).

Benoit, Kenneth, Thomas Bräuninger, and Marc Debus. 2009a. "Challenges for Estimating Policy Preferences: Announcing an Open Access Archive of Political Documents." *German Politics* 18 (3): 441–454.

Benoit, Kenneth, Michael Laver, and Slava Mikhaylov. 2009b. "Treating Words as Data with Error: Uncertainty in Text Statements of Policy Positions." *American Journal of Political Science* 53 (2): 495–513.

Benoit, Kenneth, Drew Conway, Benjamin E. Lauderdale, Michael Laver, and Slava Mikhaylov. 2016. "Crowd-Sourced Text Analysis: Reproducible and Agile Production of Political Data." *American Political Science Review* 110 (2): 278–295.

Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. 2018. "quanteda: An R Package for the Quantitative Analysis of Textual Data." *The Journal of Open Source Software* 3 (30): 774.

Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20 (3): 351–368.

Berinsky, Adam J., Michele F. Margolis, and Michael W. Sances. 2014. "Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys." *American Journal of Political Science* 58 (3): 739–753.

Bischof, Daniel, and Roman Senninger. 2018. "Simple Politics for the People? Complexity in Campaign Messages and Political Knowledge." *European Journal of Political Research* 57 (2): 473–495.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3: 993–1022.

Böhmelt, Tobias, Lawrence Ezrow, Roni Lehrer, and Hugh Ward. 2016. "Party Policy Diffusion." *American Political Science Review* 110 (2): 397–410.

Bondi, Marina, and Mike Scott, eds. 2010. *Keyness in Texts*. Amsterdam: John Benjamins.

Bovens, Mark, Thomas Schillemans, and Paul 't Hart. 2008. "Does Public Accountability Work? An Assessment Tool." *Public Administration* 86 (1): 225–242.

Bowler, Shaun, Thomas Bräuninger, Marc Debus, and Indridi H. Indridason. 2016. "Let's Just Agree to Disagree: Dispute Resolute Mechanisms in Coalition Agreements." *The Journal of Politics* 78 (4): 1264–1278.

Boydstun, Amber E. 2013. *Making the News: Politics, the Media, and Agenda Setting*. Chicago: University of Chicago Press.

Bräuninger, Thomas, and Nathalie Giger. 2018. "Strategic Ambiguity of Party Positions in Multi-Party Competition." *Political Science Research and Methods* 6 (3): 527–548.

Brooks, Deborah Jordan, and John G. Geer. 2007. "Beyond Negativity: The Effects of Incivility on the Electorate." *American Journal of Political Science* 51 (1): 1–16.

Budge, Ian. 2015. "Issue Emphases, Saliency Theory and Issue Ownership: A Historical and Conceptual Analysis." *West European Politics* 38 (4): 761–777.

Budge, Ian, and Dennis J. Farlie. 1977. *Voting and Party Competition*. London: Wiley.

Budge, Ian, and Thomas Meyer. 2013. "Understanding and Validating the Left-Right Scale (RILE)." *Mapping Policy Preferences From Texts: Statistical Solutions For*

Manifesto Analysts. Ed. by Andrea Volkens, Judith Bara, Ian Budge, Michael D. McDonald, and Hans-Dieter Klingemann. Oxford: Oxford University Press: 85–106.

Budge, Ian, Hans-Dieter Klingemann, Andrea Volkens, Judith Bara, and Eric Tanenbaum, eds. 2001. *Mapping Policy Preferences: Estimates for Parties, Electors, and Governments 1945-1998*. Oxford: Oxford University Press.

Burlacu, Diana, Ellen M. Immergut, Maria Oskarson, and Björn Rönnerstrand. 2018. "The Politics of Credit Claiming: Rights and Recognition in Health Policy Feedback." *Social Policy & Administration* 52 (4): 880–894.

Cacioppo, John T., and Wendi L. Garnder. 1999. "Emotion." *Annual Review of Psychology* 50: 191–214.

Campbell, James E. 1991. "The Presidential Surge and its Midterm Decline in Congressional Elections, 1868–1988." *The Journal of Politics* 53 (2): 477–487.

Chandler, Jesse, and Danielle Shapiro. 2016. "Conducting Clinical Research Using Crowdsourced Convenience Samples." *Annual Review of Clinical Psychology* 12: 53–81.

Chong, Dennis, and James N. Druckman. 2007. "Framing Theory." *Annual Review of Political Science* 10: 103–126.

Clark, Michael. 2014. "Understanding Parties' Policy Shifts in Western Europe: The Role of Valence 1976–2003." *British Journal of Political Science* 44 (2): 261–286.

Clarke, Nick, Will Jennings, Jonathan Moss, and Gerry Stoker. 2018. *The Good Politician: Folk Theories, Political Interaction, and the Rise of Anti-Politics*. Cambridge: Cambridge University Press.

Coppock, Alexander, Emily Ekins, and David Kirby. 2018. "The Long-lasting Effects of Newspaper Op-Eds on Public Opinion." *Quarterly Journal of Political Science* 13 (1): 59–87.

Crabtree, Charles, Matt Golder, Thomas Gschwend, and Indridi H. Indridason. 2018. *It's Not Only What you Say, It's Also How You Say It: The Strategic Use of Campaign Sentiment*. SocArXiv. URL: https://10.31235/osf.io/g2sd6.

Crepaz, Michele, and Raj S. Chari. 2018. "Assessing the Validity and Reliability of Measurements when Evaluating Public Policy." *Journal of Public Policy* 38 (3): 275–304.

Curini, Luigi. 2015. "The Conditional Ideological Inducement to Campaign on Character Valence Issues in Multiparty Systems: The Case of Corruption." *Comparative Political Studies* 48 (2): 168–192.

Cuzán, Alfred G. 2015. "Five Laws of Politics." *PS: Political Science & Politics* 48 (3): 415–419.

Dahlberg, Stefan. 2013. "Does Context Matter – The Impact of Electoral Systems, Political Parties and Individual Characteristics on Voters' Perceptions of Party Positions." *Electoral Studies* 32 (4): 670–683.

Dalton, Russell J., and Martin P. Wattenberg, eds. 2000. *Parties without Partisans: Political Change in Advanced Industrial Democracies*. Oxford: Oxford University Press.

Dalton, Russell J., David M. Farrell, and Ian McAllister. 2011. *Political Parties and the Democratic Linkage: How Parties Organize Democracy*. New York: Oxford University Press.

Damstra, Alyt, and Mark Boukes. 2018. "The Economy, the News, and the Public: A Longitudinal Study of the Impact of Economic News on Economic Evaluations and Expectations." *Communication Research* published ahead of print (doi: 10.1177/0093650217750971).

Däubler, Thomas. 2012. "The Preparation and Use of Election Manifestos: Learning from the Irish Case." *Irish Political Studies* 27 (1): 51–70.

Däubler, Thomas, Kenneth Benoit, Slava Mikhaylov, and Michael Laver. 2012. "Natural Sentences as Valid Units for Coded Political Texts." *British Journal of Political Science* 41 (4): 937–951.

De Beus, Jos. 2011. "Audience Democracy: An Emerging Pattern in Postmodern Political Communication." *Political Communication in Postmodern Democracy: Challenging the Primacy of Politics*. Ed. by Kees Brants, and Katrin Voltmer. Houndmills: Palgrave Macmillan: 19–38.

De Vreese, Claes H., Susan A. Banducci, Holli A. Semetko, and Hajo G. Boomgaarden. 2006. "The News Coverage of the 2004 European Parliamentary Election Campaign in 25 Countries." *European Union Politics* 7 (4): 477–504.

De Vries, Catherine E., Erica E. Edwards, and Erik R. Tillman. 2011. "Clarity of Responsibility Beyond the Pocketbook: How Political Institutions Condition EU Issue Voting." *Comparative Political Studies* 44 (3): 339–363.

De Vries, Erik, Martijn Schoonvelde, and Gijs Schumacher. 2018. "No Longer Lost in Translation: Evidence that Google Translate Works for Comparative Bag-of-Words Text Applications." *Political Analysis* 26 (4): 417–430.

Dodds, Peter Sheridan, Eric M. Clark, Suma Desu, Morgan R. Frank, Andrew J. Reagan, Jake Ryland Williams, Lewis Mitchell, Kameron Decker Harris, Isabel M. Kloumann, James P. Bagrow, Karine Megerdoomian, Matthew T. McMahon, Brian F. Tivnan, and Christopher M. Danforth. 2015. "Human Language Reveals a Universial Positivity Bias." *Proceedings of the National Academy of Sciences of the United States of America (PNAS)* 112 (8): 2389–2394.

Dolezal, Martin, Laurenz Ennser-Jedenastik, Wolfgang C. Müller, and Anna Katharina Winkler. 2012. "The Life Cycle of Party Manifestos: The Austrian Case." *West European Politics* 35 (4): 869–896.

Dolezal, Martin, Laurenz Ennser-Jedenastik, Wolfgang C. Müller, and Anna Katharina Winkler. 2016. "Analyzing Manifestos in their Electoral Context: A New

Approach Applied to Austria, 2002–2008." *Political Science Research and Methods* 4 (3): 641–650.

Dolezal, Martin, Laurenz Ennser-Jedenastik, Wolfgang C. Müller, Katrin Praprotnik, and Anna Katharina Winkler. 2018. "Beyond Salience and Position Taking: How Political Parties Communicate Through Their Manifestos." *Party Politics* 24 (3): 240–252.

D'Orazio, Vito, Michael Kenwick, Matthew Lane, Glenn Palmer, and David Reitter. 2016. "Crowdsourcing the Measurement of Interstate Conflict." *PLoS One* 11 (6): 1–21.

Döring, Holger. 2016. "Mapping Established Democracies: Integrated Data on Parties, Elections and Cabinets." *Electoral Studies* 44: 535–543.

Döring, Holger, and Philip Manow. 2018. *Parliaments and Governments Database (ParlGov): Information on Parties, Elections and Cabinets in Modern Democracies. Development Version*. URL: http://parlgov.org.

Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper & Row.

Duch, Raymond M., and Randolph T. Stevenson. 2008. *The Economic Vote: How Political and Economic Institutions Condition Election Results*. Cambridge: Cambridge University Press.

Dunaway, Johanna. 2013. "Media Ownership and Story Tone in Campaign News." *American Politics Research* 41 (1): 24–53.

Dupont, Julia C., Evelyn Bytzek, Melanie C. Steffens, and Frank M. Schneider. 2019. "Which Kind of Political Campaign Messages Do People Perceive as Election Pledges?." *Electoral Studies* 57: 121–130.

Duval, Dominic, and François Pétry. 2018a. "Citizens' Evaluations of Campaign Pledge Fulfillment in Canada." *Party Politics* published ahead of print (doi: 10.1177/1354068818789968).

Duval, Dominic, and François Pétry. 2018b. "Time and the Fulfillment of Election Pledges." *Political Studies* published ahead of print (doi: 10.1177/0032321718762882).

Ecker, Alejandro, and Thomas M. Meyer. 2017. "Coalition Bargaining Duration in Multiparty Democracies." *British Journal of Political Science* published ahead of print (doi: 10.1017/S0007123417000539).

Eder, Nikolaus, Marcelo Jenny, and Wolfgang C. Müller. 2017. "Manifesto Functions: How Party Candidates View and Use Their Party's Central Policy Document." *Electoral Studies* 45: 75–87.

Eichorst, Jason, and Nick Lin. 2019. "Resist to Commit: Concrete Campaign Statements and the Need to Clarify a Partisan Reputation." *The Journal of Politics* 81 (1): 15–32.

Elmelund-Præstekær, Christian. 2010. "Beyond American Negativity: Toward a General Understanding of the Determinants of Negative Campaigning." *European Political Science Review* 2 (1): 137–156.

Ennser-Jedenastik, Laurenz, and Thomas M. Meyer. 2018. "The Impact of Party Cues on Manual Coding of Political Text." *Political Science Research and Methods* 6 (3): 625–633.

Erikson, Robert S. 1988. "The Puzzle of Midterm Loss." *The Journal of Politics* 50 (4): 1011–1029.

Eshbaugh-Soha, Matthew. 2010. "The Tone of Local Presidential News Coverage." *Political Communication* 27 (1): 121–140.

Fernandez-Vazquez, Pablo. 2014. "And Yet It Moves: The Effect of Election Platforms on Party Policy Images." *Comparative Political Studies* 47 (14): 1919–1944.

Fianna Fáil. 2002. *A Lot Done: More To Do. Manifesto 2002–2007*. Dublin.

Fieldhouse, Edward, Jane Green, Geoffrey Evans, Herrmann Schmitt, Cees Van der Eijk, Jonathan Mellon, and Christopher Prosser. 2018. *British Election Study, 2017: Face-to-Face Survey.*

Folke, Olle, and James M., Jr. Snyder. 2012. "Gubernatorial Midterm Slumps." *American Journal of Political Science* 56 (4): 931–948.

Fournier, Patrick, Fred Cutler, Stuart N. Soroka, and Dietlind Stolle. 2015. *The 2015 Canadian Election Study.*

Geer, John G. 2012. "The News Media and the Rise of Negativity in Presidential Campaigns." *PS: Political Science & Politics* 45 (3): 422–427.

Gerring, John. 2012. *Social Science Methodology: A Unified Framework.* 2nd edition. Cambridge: Cambridge University Press.

Gilardi, Fabrizio, Charles R. Shipan, and Bruno Wueest. 2018. *Policy Diffusion: The Issue-Definition Stage.* Unpublished Manuscript. URL: https://fabriziogilardi.org/resources/papers/diffusion-policy-perceptions.pdf.

Goldberg, Yoav. 2016. "A Primer on Neural Network Models for Natural Language Processing." *Journal of Artificial Intelligence Research* 57: 345–420.

Goplerud, Max, and Petra Schleiter. 2016. "An Index of Assembly Dissolution Powers." *Comparative Political Studies* 49 (4): 427–456.

Green, Jane, and Will Jennings. 2017. *The Politics of Competence: Parties, Public Opinion and Voters.* Cambridge: Cambridge University Press.

Greene, Zachary. 2016. "Competing on the Issues: How Experience in Government and Economic Conditions Influence the Scope of Parties' Policy Preferences." *Party Politics* 22 (6): 809–822.

Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21 (3): 267–297.

Grimmer, Justin, Solomon Messing, and Sean J. Westwood. 2012. "How Words and Money Cultivate a Personal Vote: The Effect of Legislator Credit Claiming on Constituent Credit Allocation." *American Political Science Review* 106 (4): 703–719.

Gross, Martin, and Marc Debus. 2018. "Does EU Regional Policy Increase Parties' Support for European Integration?." *West European Politics* 41 (3): 594–614.

Grossback, Lawrence J., David A.M. Peterson, and James A. Stimson. 2005. "Comparing Competing Theories on the Causes of Mandate Perceptions." *American Journal of Political Science* 49 (2): 406–419.

Hainmueller, Jens, Dominik Hangartner, and Teppei Yamamoto. 2015. "Validating Vignette and Conjoint Survey Experiments against Real-World Behavior." *Proceedings of the National Academy of Sciences of the United States of America (PNAS)* 112 (8): 2395–2400.

Håkansson, Nicklas, and Elin Naurin. 2016. "Promising Ever More: An Empirical Account of Swedish Parties' Pledge Making During 20 Years." *Party Politics* 22 (3): 393–404.

Hallin, Daniel C., and Paolo Mancini. 2004. *Comparing Media Systems: Three Models of Media and Politics*. Cambridge: Cambridge University Press.

Harder, Raymond A., Julie Sevenans, and Peter Van Aelst. 2017. "Intermedia Agenda Setting in the Social Media Age: How Traditional Players Dominate the News Agenda in Election Times." *The International Journal of Press/Politics* 22 (3): 275–293.

Harmel, Robert. 2018. "The How's and Why's of Party Manifestos: Some Guidance for a Cross-National Research Agenda." *Party Politics* 24 (3): 229–239.

Harmel, Robert, and Kenneth Janda. 1994. "An Integrated Theory of Party Goals and Party Change." *Journal of Theoretical Politics* 6 (3): 259–287.

Haselmayer, Martin, and Marcelo Jenny. 2017. "Sentiment Analysis of Political Communication: Combining a Dictionary Approach with Crowdcoding." *Quality & Quantity* 51 (6): 2623–2646.

Healy, Andrew, and Neil Malhotra. 2013. "Retrospective Voting Reconsidered." *Annual Review of Political Science* 16: 285–306.

Hitt, Matthes P., and Kathleen Searles. 2018. "Media Coverage and Public Approval of the U.S. Supreme Court." *Political Communication* 35 (4): 566–586.

Hix, Simon, and Michael Marsh. 2007. "Punishment or Protest? Understanding European Parliament Elections." *The Journal of Politics* 69 (2): 495–501.

Hix, Simon, and Michael Marsh. 2011. "Second-Order Effects Plus Pan-European Political Swings: An Analysis of European Parliament Elections Across Time." *Electoral Studies* 30: 4–15.

Hopkins, Daniel J., and Gary King. 2010. "A Method of Automated Nonparametric Content Analysis for Social Science." *American Journal of Political Science* 54 (1): 229–247.

Hopmann, David Nicholas, Rens Vliegenthart, and Jürgen Maier. 2018. "The Effects of Tone, Focus, and Incivility in Election Debates." *Journal of Elections, Public Opinion and Parties* 28 (3): 283–306.

Horiuchi, Yusaku, Daniel M. Smith, and Teppei Yamamoto. 2018. "Measuring Voters' Multidimensional Policy Preferences with Conjoint Analysis: Application to Japan's 2014 Election." *Political Analysis* 26 (2): 190–209.

Horn, Alexander. 2018. "Can the Online Crowd Match Real Expert Judgments? How Task Complexity and Coder Location Affect the Validty of Crowd-Coded Data." *European Journal of Political Research* 598 (1): 236–247.

Horn, Alexander, and Carsten Jensen. 2017. "When and Why Politicians Do Not Keep Their Welfare Promises." *European Journal of Political Research* 56 (2): 381–400.

Huff, Connor, and Joshua D. Kertzer. 2018. "How The Public Defines Terrorism." *American Journal of Political Science* 62 (1): 55–71.

ISSP Research Group. 2008. *International Social Survey Programme: Role of Government IV – ISSP 2006*. GESIS Data Archive, Cologne. ZA4700 Data file Version 1.0.0. URL: https://10.4232/1.4700.

ISSP Research Group. 2018. *International Social Survey Programme: Role of Government IV – ISSP 2016*. GESIS Data Archive, Cologne. ZA6900 Data file Version 1.0.0. URL: http://doi.org/10.4232/1.12994.

Iyengar, Shanto. 1991. *Is Anyone Responsible?: How Television Frames Political Issues*. Chicago: University of Chicago Press.

Jeffery, Charlie, and Daniel Hough. 2001. "The Electoral Cycle and Multi-Level Voting in Germany." *German Politics* 10 (2): 73–98.

Jennings, Will, and Christopher Wlezien. 2016. "The Timeline of Elections: A Comparative Perspective." *American Journal of Political Science* 60 (1): 219–233.

Jennings, Will, and Christopher Wlezien. 2018. "Election Polling Errors Across Time and Space." *Nature Human Behaviour* 2 (4): 276–283.

Johns, Robert. 2011. "Credit Where it's Due? Valence Politics, Attributions of Responsibility, and Multi-Level Elections." *Political Behavior* 33 (1): 53–77.

Kalogeropoulou, Efthalia. 1989. "Election Promises and Government Performance in Greece: PASOK's Fulfilment of its 1981 Election Pledges." *European Journal of Political Research* 17 (3): 289–311.

Kang, Woo Chang. 2015. "Electoral Cycles in Pork Barrell Politics: Evidence from South Corea 1989–2008." *Electoral Studies* 38: 46–58.

Kern, Holger Lutz, and Jens Hainmueller. 2006. "Electoral Balancing, Divided Government and 'Midterm' Loss in German Elections." *Journal of Legislative Studies* 12 (2): 127–149.

King, Gary, Patricka Lam, and Margaret E. Roberts. 2017. "Computer-Assisted Keyword and Document Set Discovery from Unstructured Text." *American Journal of Political Science* 61 (4): 971–988.

Klingemann, Hans-Dieter, Richard I. Hofferbert, and Ian Budge. 1994. *Parties, Policies, and Democracy*. Boulder: Westview Press.

Klüver, Heike, and Radoslaw Zubek. 2018. "Minority Governments and Legislative Reliability: Evidence from Denmark and Sweden." *Party Politics* 24 (6): 719–730.

Kosmidis, Spyros, Sara B. Hobolt, Eamonn Molloy, and Stephen Whitefield. 2018. "Party Competition and Emotive Rhetoric." *Comparative Political Studies* published ahead of print (doi: 10.1177/0010414018797942).

Kostadinova, Petia. 2017. "Party Pledges in the News: Which Election Promises Do the Media Report?." *Party Politics* 23 (6): 636–645.

Krippendorff, Klaus. 2004. "Reliability in Content Analysis: Some Common Misconceptions and Recommendations." *Human Communication Research* 30 (3): 411–433.

Ladd, Jonathan McDonald, and Gabriel S. Lenz. 2009. "Exploiting a Rare Communication Shift to Document the Persuasive Power of the News Media." *American Journal of Political Science* 53 (2): 394–410.

Langer, Ana Ines, and Iñaki Sagarzazu. 2018. "Bring Back the Party: Personalisation, the Media and Coalition Politics." *West European Politics* 41 (2): 472–495.

Lau, Richard R., and Gerald M. Pomper. 2002. "Effectiveness of Negative Campaigning in U.S. Senate Elections." *American Journal of Political Science* 46 (1): 47–66.

Laver, Michael. 2011. "Why Vote-Seeking Parties May Make Voters Miserable." *Irish Political Studies* 26 (4): 489–500.

Laver, Michael. 2014. "Measuring Policy Positions in Political Space." *Annual Review of Political Science* 17: 207–223.

Laver, Michael, John Garry, and Kenneth Benoit. 2003. "Extracting Policy Positions from Political Texts Using Words as Data." *American Political Science Review* 97 (2): 311–331.

Lazer, David, and Jason Radford. 2017. "Data ex Machina: Introduction to Big Data." *Annual Review of Sociology* 43: 19–39.

Lehmann, Pola, and Malisa Zobel. 2018. "Positions and Saliency of Immigration in Party Manifestos: A Novel Dataset Using Crowd Coding." *European Journal of Political Research* 57 (4): 1056–1083.

Lehmann, Pola, Theres Matthieß, Nicolas Merz, Sven Regel, and Annika Werner. 2017. *Manifesto Corpus: Version 2017-2*. Berlin: WZB Berlin Social Science Center.

Lengauer, Günther, Frank Esser, and Rosa Berganza. 2011. "Negativity in Political News: A Review of Concepts, Operationalizations and Key Findings." *Journalism* 13 (2): 179–202.

Lewis-Beck, Michael S., and Mary Stegmaier. 2000. "Economic Determinants of Electoral Outcomes." *Annual Review of Political Science* 3: 183–219.

Lijphart, Arend. 2012. *Patterns of Democracy: Government Forms and Performance in Thirty-Six Countries*. 2nd edition. New Haven: Yale University Press.

Lind, Fabienne, Maria Gruber, and Hajo G. Boomgaarden. 2017. "Content Analysis by the Crowd: Assessing the Usability of Crowdsourcing for Coding Latent Constructs." *Communication Methods and Measures* 11 (3): 191–209.

Linder, Fridolin, Bruce A. Desmarais, Matthew Burgess, and Eugenia Giraudy. 2018. "Text as Policy: Measuring Policy Similarity through Bill Text Reuse." *Policy Studies Journal* published ahead of print (doi: 10.1111/psj.12257).

Lindgren, Elina. 2018. "Changing Policy With Words: How Persuasive Words in Election Pledges Influence Voters' Beliefs About Policies." *Mass Communication and Society* 21 (4): 425–449.

Lo, James, Sven-Oliver Proksch, and Jonathan B. Slapin. 2016. "Ideological Clarity in Multiparty Competition: A New Measure and Test Using Election Manifestos." *British Journal of Political Science* 46 (3): 591–610.

Lowe, Will, and Kenneth Benoit. 2013. "Validating Estimates of Latent Traits from Textual Data Using Human Judgment as a Benchmark." *Political Analysis* 21 (3): 298–313.

Lowe, Will, Kenneth Benoit, Slava Mikhaylov, and Michael Laver. 2011. "Scaling Policy Preferences from Coded Political Texts." *Legislative Studies Quarterly* 36 (1): 125–155.

Lucas, Christopher, Richard A. Nielsen, Margaret E. Roberts, Brandon M. Stewart, Alex Storer, and Dustin Tingley. 2015. "Computer-Assisted Text Analysis for Comparative Politics." *Political Analysis* 23 (2): 254–277.

Maier, Jürgen, and Carolin Jansen. 2017. "When Do Candidates Attack in Election Campaigns? Exploring the Determinants of Negative Candidate Messages in German Televised Debates." *Party Politics* 34 (5): 549–559.

Maier, Jürgen, and Anna-Maria Renner. 2018. "When a Man Meets a Woman: Comparing the Use of Negativity of Male Candidates in Single- and Mixed-Gender Televised Debates." *Political Communication* 35 (3): 433–449.

Mair, Peter. 2008. "The Challenge to Party Government." *West European Politics* 31 (1–2): 211–234.

Mair, Peter. 2009. *Representative versus Responsible Government*. MPIfG Working Paper. URL: http://edoc.vifapol.de/opus/volltexte/2010/2121/.

Mair, Peter. 2013. *Ruling the Void: The Hollowing of Western Democracy*. London: Verso.

Manin, Bernard, Adam Przeworski, and Susan C. Stokes. 1999. "Elections and Representation." *Democracy, Accountability, and Representation*. Ed. by Adam

Przeworski, Susan C. Stokes, and Bernard Manin. Cambridge: Cambridge University Press: 29–54.

Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schütze. 2008. *An Introduction to Information Retrieval*. New York: Cambridge University Press.

Mansbridge, Jane. 2003. "Rethinking Representation." *American Political Science Review* 97 (4): 515–528.

Marquardt, Kyle L., Daniel Pemstein, Constanza Sanhueza Petrarca, Brigitte Seim, Steven Lloyd Wilson, Michael Bernhard, Michael Coppedge, and Staffan I. Lindberg. 2017. *Experts, Coders, and Crowds: An Analysis of Substitutability*. The Varieties of Democracy Institute Working Paper Series, 53. URL: https://www.v-dem.net/media/filer_public/9e/81/9e81b209-4b20-4188-878b-9ed92781ff56/v-dem_working_paper_2017_53.pdf.

Marsh, Michael, and James Tilley. 2010. "The Attribution of Credit and Blame to Governments and Its Impact on Vote Choice." *British Journal of Political Science* 40 (1): 115–134.

Martin, Lanny W., and Georg Vanberg. 2003. "Wasting Time? The Impact of Ideology and Size on Delay in Coalition Formation." *British Journal of Political Science* 33 (2): 323–332.

McAllister, Ian, Clive Bean, Rachel Gibson, Toni Makkai, and Jill Sheppard. 2016. *Australian Election Study 2016*.

McCombs, Maxwell E., and Donald L. Shaw. 1972. "The Agenda-Setting Function of Mass Media." *Public Opinion Quarterly* 36 (2): 176–187.

McDonald, Michael D., and Ian Budge. 2005. *Elections, Parties, Democracy: Conferring the Median Mandate*. Oxford: Oxford University Press.

Merz, Nicolas. 2017. "Gaining Voice in the Mass Media: The Effect of Parties' Strategies on Party-Issue Linkages in Election News Coverage." *Acta Politica* 52 (4): 436–460.

Merz, Nicolas, Sven Regel, and Jirka Lewandowski. 2016. "The Manifesto Corpus: A New Resource for Research on Political Parties and Quantitative Text Analysis." *Research and Politics* 3 (2): 1–8.

Michelitch, Kristin, and Stephen Utych. 2018. "Electoral Cycle Fluctuations in Partisanship: Global Evidence from Eighty-Six Countries." *The Journal of Politics* 80 (2): 412–427.

Mikhaylov, Slava, Michael Laver, and Kenneth Benoit. 2012. "Coder Reliability and Misclassifcation in the Human Coding of Party Manifestos." *Political Analysis* 20 (1): 78–91.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. *Efficient Estimation of Word Representations in Vector Space*. arXiv: 1301.3781. URL: https://arxiv.org/abs/1301.3781.

Miller, W. L., and M. Mackie. 1973. "The Electoral Cycle and the Asymmetry of Government and Opposition Popularity: An Alternative Model of the Relationship Between Economic Conditions and Political Popularity." *Political Studies* 21 (3): 263–279.

Moury, Catherine, and Jorge M. Fernandes. 2018. "Minority Governments and Pledge Fulfilment: Evidence from Portugal." *Government and Opposition* 53 (2): 335–355.

Muddiman, Ashley, and Natalie Jomini Stroud. 2017. "News Values, Cognitive Biases, and Partisan Incivility in Comment Sections." *Journal of Communication* 67 (4): 586–609.

Muddiman, Ashley, Shannon C. McGregor, and Natalie Jomini Stroud. 2018. "(Re)Claiming Our Expertise: Parsing Large Text Corpora With Manually Validated and Organic Dictionaries." *Political Communication* published ahead of print (doi: 10.1080/10584609.2018.1517843).

Mueller, Hannes, and Christopher Rauh. 2018. "Reading Between the Lines: Prediction of Political Violence Using Newspaper Text." *American Political Science Review* 112 (2): 358–375.

Mueller, John E. 1970. "Presidential Popularity from Truman to Johnson." *American Political Science Review* 64 (1): 18–34.

Müller, Stefan, and Tom Louwerse. 2018. "The Electoral Cycle Effect in Parliamentary Democracies." *Political Science Research and Methods* published ahead of print (doi: 10.1017/psrm.2018.47).

Müller, Wolfgang C., and Kaare Strøm, eds. 1999. *Policy, Office, or Votes? How Political Parties in Western Europe Make Hard Decisions*. Cambridge: Cambridge University Press.

Müller, Wolfgang C., and Kaare Strøm. 2008. "Coalition Agreements and Cabinet Governance." *Cabinet Governance: Bargaining and the Cycle of Democratic Politics*. Ed. by Kaare Strøm, Wolfgang C. Müller, and Torbjörn Bergman. Oxford University Press: 159–199.

Müller, Wolfgang C., Anita Bodlos, Martin Dolezal, Nikolaus Eder, Laurenz Ennser-Jedenastik, Matthias Kaltenegger, Thomas M. Meyer, Katrin Praprotnik, and Anna Katharina Winkler. 2017. *AUTNES Content Analysis of Party Manifestos 2013: ZA6877 Data File Version 1.1.0*. GESIS Data Archive, Cologne. URL: http://doi.org/10.4232/1.12752.

Mullinix, Kevin J., Thomas J. Leeper, James N. Druckman, and Jeremy Freese. 2015. "The Generalizability of Survey Experiments." *Journal of Experimental Political Science* 2 (2): 109–138.

Nannestad, Peter, and Martin Paldam. 1999. "The Cost of Ruling: A Foundation Stone of Two Theories." University of Aarhus Department of Economics, Working Paper No. 1999–9.

Naurin, Elin. 2011. *Election Promises, Party Behaviour and Voter Perceptions*. Houndmills: Palgrave Macmillan.

Naurin, Elin. 2014. "Is Promise a Promise? Election Pledge Fulfilment in Comparative Perspective Using Sweden as an Example." *West European Politics* 37 (5): 1046–1064.

Naurin, Elin. 2016. "Do They Really Know? On Party Representatives' Knowledge of the Extent to Which Their Parties Make Election Pledges." *Scandinavian Political Studies* 39 (4): 388–410.

Naurin, Elin, and Henrik Ekengren Oscarsson. 2017. "When and Why Are Voters Correct in Their Evaluations of Specific Government Performance?." *Political Studies* 65 (4): 860–876.

Naurin, Elin, Terry J. Royed, and Robert Thomson, eds. 2019. *Party Mandates and Democracy: Making, Breaking, and Keeping Election Pledges in Twelve Countries*. Ann Arbor: University of Michigan Press.

Patterson, Thomas E. 1993. *Out of Order: An Incisive and Boldly Original Critique of the News Media's Domination of America's Political Process*. New York: Vintage.

Patterson, Thomas E., and Wolfgang Donsbagh. 1996. "News Decisions: Journalists as Partisan Actors." *Political Communication* 13 (4): 455–468.

Pennebaker, James W., Ryan L. Boyd, Kayla Jordan, and Kate Blackburn. 2015. *The Development and Psychometric Properties of LIWC2015*. Austin: University of Texas at Austin.

Pérez, Efren O., and Margit Tavits. 2017. "Language Shapes People's Time Perspective and Support for Future-Oriented Policies." *American Journal of Political Science* 61 (3): 715–727.

Petrocik, John R. 1996. "Issue Ownership in Presidential Elections, with a 1980 Case Study." *American Journal of Political Science* 40 (3): 825–850.

Pétry, François, Dominic Duval, and Lisa Birch. 2018. "Do Regional Governments Fulfill Fewer Election Promises than National Governments? The Case of Quebec in Comparative Perspective." *Party Politics* published ahead of print (doi: 10.1177/1354068818787353).

Pinheiro, José, and Douglas Bates. 2000. *Mixed-Effects Models in S and S-Plus*. New York: Springer.

Pomper, Gerald M., and Susan S. Lederman. 1980. *Elections in America: Control and Influence in Democratic Politics*. 2nd edition. New York: Longman.

Powell, G. Bingham. 2000. *Elections as Instruments of Democracy: Majoritarian and Proportional Visions*. New Haven: Yale University Press.

Powell, G. Bingham, and Guy D. Whitten. 1993. "A Cross-National Analysis of Economic Voting: Taking Account of the Political Context." *American Journal of Political Science* 37 (2): 391–414.

Praprotnik, Katrin. 2017. "Jurisdiction, Time, and Money: The Role of Resources in Pledge Fulfilment." *Party Politics* 23 (6): 848–859.

Prior, Markus. 2017. "Conditions for Political Accountability in a High-Choice Media Environment." *Oxford Handbook of Political Communication*. Ed. by Kathleen Hall Jamieson, Dan Kahan, and Dietram A. Scheufele. Oxford: Oxford University Press: 897–910.

Proksch, Sven-Oliver, and Jonathan B. Slapin. 2015. *The Politics of Parliamentary Debate: Parties, Rebels, and Representation*. Cambridge: Cambridge University Press.

Proksch, Sven-Oliver, Will Lowe, Jens Wäckerle, and Stuart N. Soroka. 2019. "Multilingual Sentiment Analysis: A New Approach to Measuring Conflict in Legislative Speeches." *Legislative Studies Quarterly* 44 (1): 97–131.

Rauh, Christian. 2018. "Validating a Sentiment Dictionary for German Political Language: A Workbench Note." *Journal of Information Technology & Politics* 15 (4): 319–343.

Reck, Christian, and Reinhold Weber. 2010. "Statistischer Anhang." *Parteien in den deutschen Ländern*. Ed. by Andreas Kost, Werner Rellecke, and Reinhold Weber. Munich: C.H Beck: 410–446.

Reeves, Aaron, Martin McKee, and David Stuckler. 2016. "'It's The Sun Wot Won It': Evidence of Media Influence on Political Attitudes and Voting from a UK Quasi-natural Experiment." *Social Science Research* 56: 44–57.

Reif, Karlheinz, and Hermann Schmitt. 1980. "Nine Second-Order National Elections: A Conceptual Framework for the Analysis of European Election Results." *European Journal of Political Research* 8 (1): 3–44.

Roberts, Andrew. 2010. *The Quality of Democracy in Eastern Europe: Public Preferences and Policy Reforms*. Cambridge: Cambridge University Press.

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58 (4): 1064–1082.

Roberts, Marilyn, and Maxwell E. McCombs. 1994. "Agenda Setting and Political Advertising: Origins of the News Agenda." *Political Communication* 11 (3): 249–262.

Rose, Richard. 1984. *Do Parties Make a Difference?* 2nd edition. London: Macmillan.

Rovny, Jan. 2012. "Who Emphasizes and Who Blurs? Party Strategies in Multidimensional Competition." *European Union Politics* 13 (2): 269–292.

Royed, Terry J. 1996. "Testing the Mandate Model in Britain and the United States: Evidence from the Reagan and Thatcher Eras." *British Journal of Political Science* 26 (1): 45–80.

Royed, Terry J., Elin Naurin, and Robert Thomson. 2019. "A Systematic Research Design for Studying the Fulfillment of Election Pledges." *Party Mandates and Democracy: Making, Breaking, and Keeping Election Pledges in Twelve Countries*. Ed. by Elin Naurin, Terry J. Royed, and Robert Thomson. Ann Arbor: University of Michigan Press: 23–28.

Rozin, Paul, and Edward B. Royzman. 2001. "Negativity Bias, Negativity Dominance, and Contagion." *Personality and Social Psychology Review* 5 (4): 296–320.

Rudolph, Thomas J. 2003. "Who's Responsible for the Economy? The Formation and Consequences of Responsibility Attributions." *American Journal of Political Science* 47 (4): 698–713.

Sagarzazu, Iñaki, and Heike Klüver. 2017. "Coalition Governments and Party Competition: Political Communication Strategies of Coalition Parties." *Political Science Research and Methods* 5 (2): 333–349.

Salganik, Matthew J. 2017. *Bit by Bit: Social Research in the Digital Age*. Princeton: Princeton University Press.

Sartori, Giovanni. 1976. *Parties and Party Systems: A Framework for Analysis*. New York: Cambridge University Press.

Schleiter, Petra, and Margit Tavits. 2016. "The Electoral Benefits of Opportunistic Election Timing." *The Journal of Politics* 78 (3): 836–850.

Schmitt, Herrmann, and Eftichia Teperoglou. 2017. "The Study of Less Important Elections." *The SAGE Handbook of Electoral Behavior*. Ed. by Kai Arzheimer, Jocelyn Evans, and Michael S. Lewis-Beck. London: SAGE: 56–79.

Schroeder, Alan. 2008. *Presidential Debates: Fifty Years of High-Risk TV*. New York: Columbia University Press.

Schumacher, Gijs, Marc Van de Wardt, Barbara Vis, and Michael Baggesen Klitgaard. 2015. "How Aspiration to Office Conditions the Impact of Government Participation

on Party Platform Change." *American Journal of Political Science* 59 (4): 1040–1054.

Seeberg, Henrik Bech. 2017. "How Stable Is Political Parties' Issue Ownership? A Cross-Time, Cross-National Analysis." *Political Studies* 65 (2): 475–492.

Serra, Gilles. 2010. "Polarization of What? A Model of Elections with Endogenous Valence." *The Journal of Politics* 72 (2): 426–437.

Shepsle, Kenneth A., Robert P. Van Houweling, Samuel Abrams, and Peter C. Hanson. 2009. "The Senate Electoral Cycle and Bicameral Appropriations Politics." *American Journal of Political Science* 53 (2): 343–359.

Shoemaker, Pamela J., and Tim P. Vos. 2009. *Gatekeeping Theory*. New York: Routledge.

Slapin, Jonathan B., and Sven-Oliver Proksch. 2008. "A Scaling Model for Estimating Time-Series Party Positions from Texts." *American Journal of Political Science* 52 (3): 705–722.

Snyder, James M., Jr., and David Strömberg. 2010. "Press Coverage and Political Accountability." *Journal of Political Economy* 118 (2): 355–408.

Snyder, Richard. 2001. "Scaling Down: The Subnational Comparative Method." *Studies in Comparative International Development* 36 (1): 93–110.

Somer-Topcu, Zeynep. 2015. "Everything to Everyone: The Electoral Consequences of the Broad-Appeal Strategy in Europe." *American Journal of Political Science* 59 (4): 841–854.

Soroka, Stuart N. 2006. "Good News and Bad News: Asymmetric Responses to Economic Information." *The Journal of Politics* 68 (2): 372–385.

Soroka, Stuart N. 2012. "The Gatekeeping Function: Distributions of Information in Media and the Real World." *The Journal of Politics* 74 (2): 514–528.

Soroka, Stuart N., and Stephen McAdams. 2015. "News, Politics, and Negativity." *Political Communication* 32 (1): 1–22.

Soroka, Stuart N., and Christopher Wlezien. 2010. *Degrees of Democracy: Politics, Public Opinion, and Policy*. Cambridge: Cambridge University Press.

Soroka, Stuart N., and Christopher Wlezien. 2018. "Tracking the Coverage of Public Policy in Mass Media." *Policy Studies Journal* published ahead of print (doi: 10.1111/psj.12285).

Soroka, Stuart N., Lori Young, and Meital Balmas. 2015. "Bad News or Mad News? Sentiment Scoring of Negativity, Fear, and Anger in News Content." *The ANNALS of the American Academy of Political and Social Science* 659 (1): 108–121.

SPD Bundestagsfraktion. 2017. *Gesagt, getan, gerecht: Bilanz 2013–2017*. URL: https://www.spdfraktion.de/system/files/documents/spdbf-bilanz-a5-2013-2017-web.pdf.

Spoon, Jae-Jae, and Heike Klüver. 2017. "Does Anybody Notice? How Policy Positions of Coalition Parties Are Perceived by Voters." *European Journal of Political Research* 56 (1): 115–132.

Stevenson, Randolph T. 2002. "The Cost of Ruling, Cabinet Duration, and the 'Median-Gap' Model." *Public Choice* 113 (1): 157–178.

Stimson, James A. 1976. "Public Support for American Presidents: A Cyclical Model." *Public Opinion Quarterly* 40 (1): 1–21.

Stimson, James A., Michael B. Mackuen, and Robert S. Erikson. 1995. "Dynamic Representation." *American Political Science Review* 89 (3): 543–565.

Stokes, Donald E. 1963. "Spatial Models of Party Competition." *American Political Science Review* 57 (2): 368–377.

Strøm, Kaare. 2000. "Delegation and Accountability in Parliamentary Democracies." *European Journal of Political Research* 37 (3): 261–289.

Strömbäck, Jesper. 2005. "In Search of a Standard: Four Models of Democracy and their Normative Implications for Journalism." *Journalism Studies* 6 (3): 331–345.

Strömbäck, Jesper. 2008. "Four Phases of Mediatization: An Analysis of the Mediatization of Politics." *The International Journal of Press/Politics* 13 (3): 228–246.

Tausczik, Yla R., and James W. Pennebaker. 2010. "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods." *Journal of Language and Social Psychology* 29 (1): 24–54.

Thesen, Gunnar. 2013. "When Good News is Scarce and Bad News is Good: Government Responsibilities and Opposition Possibilities in Political Agenda-Setting." *European Journal of Political Research* 52 (3): 364–389.

Thomson, Robert. 2001. "The Programme to Policy Linkage: The Fulfilment of Election Pledges on Socio-Economic Policy in The Netherlands, 1986–1998." *European Journal of Political Research* 40 (2): 171–197.

Thomson, Robert. 2011. "Citizens' Evaluations of the Fufillment of Election Pledges: Evidence from Ireland." *The Journal of Politics* 73 (1): 187–201.

Thomson, Robert, and Heinz Brandenburg. 2018. "Trust and Citizens' Evaluations of Promise Keeping by Government Parties." *Political Studies* published ahead of print (doi: 10.1177/0032321718764177).

Thomson, Robert, and Rory Costello. 2016. "Governing Together in Good and Bad Economic Times: The Fulfilment of Election Pledges in Ireland." *Irish Political Studies* 31 (2): 182–203.

Thomson, Robert, Terry J. Royed, Elin Naurin, Joaquín Artés, Rory Costello, Laurenz Ennser-Jedenastik, Mark Ferguson, Petia Kostadinova, Catherine Moury, François Pétry, and Katrin Praprotnik. 2017. "The Fulfillment of Parties' Election Pledges: A Comparative Study on the Impact of Power Sharing." *American Journal of Political Science* 61 (3): 527–542.

Traber, Denise, Martijn Schoonvelde, and Gijs Schumacher. Forthcoming. "Errors Have Been Made, Others Will be Blamed: Blame Avoidance and Blame-Shifting

in Prime Minister Speeches During the Economic Crisis in Europe." *European Journal of Political Research*.

Trussler, Marc, and Stuart N. Soroka. 2014. "Consumer Demand for Cynical and Negative News Frames." *The International Journal of Press/Politics* 19 (3): 360–379.

Tufte, Eward R. 1975. "Determinants of the Outcomes of Midterm Congressional Elections." *American Political Science Review* 69 (3): 812–826.

Volkens, Andrea, Judith Bara, Ian Budge, Michael D. McDonald, and Hans-Dieter Klingemann, eds. 2013. *Mapping Policy Preferences From Texts: Statistical Solutions for Manifesto Data Analysts*. Oxford: Oxford University Press.

Walgrave, Stefaan, Stuart N. Soroka, and Michiel Nuytemans. 2008. "The Mass Media's Political Agenda Setting Power: A Longitudinal Analysis of Media, Parliament, and Government in Beglium (1993 to 2000)." *Comparative Political Studies* 41 (6): 814–836.

Walther, Daniel, and Johan Hellström. 2019. "The Verdict in the Polls: How Government Stability is Affected by Popular Support." *West European Politics* 42 (3): 593–617.

Weaver, R. Kent. 1986. "The Politics of Blame Avoidance." *Journal of Public Policy* 6 (4): 371–398.

Wilkerson, John, and Andreu Casas. 2017. "Large-Scale Computerized Text Analysis in Political Science: Opportunities and Challenges." *Annual Review of Political Science* 20: 529–544.

Wilkerson, John, David Smith, and Nicholas Stramp. 2015. "Tracing the Flow of Policy Ideas in Legislatures: A Text Reuse Approach." *American Journal of Political Science* 59 (4): 943–956.

Wlezien, Christopher. 1995. "The Public as Thermostat: Dynamics of Preferences for Spending." *American Journal of Political Science* 39 (4): 981–1000.

Wlezien, Christopher. 2017. "Policy (Mis)Representation and the Cost of Ruling: U.S. Presidential Elections in Comparative Perspective." *Comparative Political Studies* 50 (6): 711–738.

Wring, Dominic, and David Deacon. 2010. "Patterns of Press Partisanship in the 2010 General Election." *British Politics* 5 (4): 436–454.

Wring, Dominic, and David Deacon. 2018. "A Bad Press: Newspapers." *The British General Election of 2017*. Ed. by Philip Cowley, and Dennis Kavanagh. Cham: Palgrave Macmillan: 347–384.

Young, Lori, and Stuart N. Soroka. 2012. "Affective News: The Automated Coding of Sentiment in Political Texs." *Political Communication* 29 (2): 205–231.