POLITICAL COMMUNICATION
2020, VOL. 37, NO. 5, 696–718
https://doi.org/10.1080/10584609.2020.1744779

Routledge
Taylor & Francis Group

Check for updates

Media Coverage of Campaign Promises Throughout the Electoral Cycle

Stefan Müller

School of Politics and International Relations, University College Dublin, Dublin, Ireland

ABSTRACT

Previous studies conclude that governments fulfill a large share of their campaign pledges. However, only a minority of voters believe that politicians try to keep their promises, and many voters struggle to recall the fulfillment or breaking of salient campaign pledges accurately. I argue that this disparity between the public perception and empirical evidence is influenced by the information voters receive throughout the electoral cycle. I expect that the media extensively inform readers about political promises. In addition, I posit that news outlets focus more on broken than on fulfilled promises and that the focus on broken promises has increased over time. I find strong support for these expectations based on a new text corpus of over 430,000 statements on political commitments published between 1979 and 2017 in 22 newspapers during 33 electoral cycles in Australia, Canada, Ireland, and the United Kingdom. Newspapers inform voters regularly about announced, broken, and fulfilled promises. Yet, across the four countries, newspapers report at least twice as much on broken than on fulfilled promises. Moreover, this negativity bias in reports on political promises has increased substantively. The results have implications for studying campaign promises, negative information in mass media, and the linkages between voters and parties.

KEYWORDS

campaign pledges; negativity bias; mass media; political communication; quantitative text analysis

Introduction

Previous research concludes that parties fulfill a large share of their manifesto promises (e.g., Costello & Thomson, 2008; Naurin et al., 2019a; Royed, 1996; Thomson, 2001; Thomson et al., 2017). However, survey evidence from developed democracies clearly shows that only few citizens believe that parties deliver on their promises (ISSP Research Group, 2018). Moreover, many voters make inaccurate evaluations when asked whether specific promises have been fulfilled (Belchior 2019; Duval & Pétry, 2018; Naurin & Oscarsson, 2017; Thomson, 2011; Thomson & Brandenburg, 2019). Scholars call this disparity between empirical evidence and public perceptions of the "pledge puzzle."

I argue that the contradiction between empirical evidence and the public perception is partially driven by the information voters receive before elections and during the electoral cycle. Voters do not experience government policy directly by reading manifestos, listening to parliamentary debates, or reading budget reports. Instead, they receive a filtered

CONTACT Stefan Müller stefan.mueller@ucd.ie School of Politics and International Relations, University College Dublin, Newman Building, Belfield, Dublin 4, Ireland.

Supplemental data for this article can be accessed on the publisher's website at https://doi.org/10.1080/10584609.2020.1744779

sample of this information through the media. Thus, the media play a crucial role in providing information required to evaluate parties' achievements and failures (McCombs & Shaw, 1972; Soroka & Wlezien, 2019). The way the media report on parties' ability to deliver on their campaign promises can influence citizens' assessment of government performance. As Kostadinova (2017, p. 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. Yet, if voters do not receive the required information to evaluate the performance of parties, the normatively appealing theory of promissory representation might not work in reality (Manin et al., 1999; Mansbridge, 2003).

Despite the theoretical and practical importance of the media for informing citizens about policy outputs, few studies have analyzed media coverage of election pledges (for notable exceptions see Duval, 2019; Kostadinova, 2017, 2019). Since media coverage can influence attitudes toward political issues or parties (e.g., Ladd & Lenz, 2009; Moy & Pfau, 2000; Reeves et al., 2016), a consistent focus on broken promises could contribute to the public perception of promise-breaking politicians. Moreover, experimental evidence shows that broken pledges are more important than fulfilled pledges when evaluating the performance of a government (Naurin et al., 2019b). If media reporting of promises is overly negative, the public perception of promise-breaking politicians could be reaffirmed. But do the media focus more on broken promises than fulfilled promises? To answer this question, we first require a descriptive, systematic, and comparative analysis of pledge coverage by media outlets (Althaus et al., 2011).

In this paper, I provide the most comprehensive and systematic study of media reporting of election promises to date in terms of the numbers of government cycles and countries covered. This study focuses on three unanswered questions using an original dataset of over 400,0000 newspaper articles about political promises across 33 electoral cycles in Australia, Canada, the Republic of Ireland, and the United Kingdom. First, how does media coverage of election promises vary throughout and across electoral cycles? Second, do news outlets focus more on fulfilled or broken pledges? Third, do we observe a stronger focus on broken promises in recent years?

The results enhance our understanding of accountability, election promises, and media coverage of politics in three ways. First, the finding that pledge coverage increases exponentially before elections is reassuring: the media inform voters about promises, policy alternatives, and "what is at stake" during an election. Second, the consistently lower degree of coverage on fulfilled promises relative to broken promises might at least partially explain why the majority of citizens does not believe that parties keep their promises (ISSP Research Group, 2018). This finding also speaks to the literature on negative and positive stories 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. Finally, I find that the negativity bias has increased over time, supporting previous conclusions (mostly drawn from the US-context) on journalists' incentives to focus on negative news (Geer, 2006, 2012; Ridout & Smith, 2008). This negativity bias could help to explain the "cost of ruling" (Klüver & Spoon, 2020; Müller & Louwerse, 2018; Wlezien, 2017) and the public perception that parties do not deliver on their promises. More broadly, the results offer a novel perspective on promissory representation (Mansbridge, 2003) and the mutual relationships between voters, parties, and the media.

Media and the Mandate Model of Democracy

In the classic version of the mandate model of democracy, voters choose between election platforms and select the party that comes closest to their policy preferences (e.g., Downs, 1957; Mansbridge, 2003). 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 for the upcoming legislative cycle (Manin et al., 1999). Citizens judge parties' performances by (not) voting for a party or candidate at the upcoming election. This democratic chain of delegation requires a constant flow of accurate information about parties' actions in parliament.

Previous studies usually code election manifestos to assess the degree of mandate fulfillment (e.g., Naurin et al., 2019a; Royed, 1996; Thomson, 2001; Thomson et al., 2017). These documents, published by almost all parties prior to elections, are the definitive statement of parties' policy positions and provide a "well-defined and coherent body of officially sanctioned promises" (Pétry et al., 2018, p. 3). While manifestos are suitable to assess party pledge fulfillment, it is unrealistic to assume that voters remember specific manifesto content when voting retrospectively or prospectively. Indeed, many voters struggle to accurately recall the fulfillment or breaking of salient pledges (Belchior 2019; Duval & Pétry, 2018; Naurin & Oscarsson, 2017; Thomson, 2011; Thomson & Brandenburg, 2019).[1] Citizens seem to have only limited knowledge of manifesto pledges and pledge fulfillment.

Rather than assessing the proportion of fulfilled manifesto 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. News coverage influences citizens' perception of political topics and problems (e.g., Chong & Druckman, 2007; Walgrave et al., 2008). Media also take a crucial gatekeeping function because journalists decide what decisions or opinions are newsworthy (Shoemaker & Vos, 2009; Soroka, 2012). Journalists function as partisan actors when determining what (not) to cover in newspaper articles (Patterson & Donsbagh, 1996), and citizens receive information on political decisions and outcomes through the media, not directly through parties and politicians (Soroka & Wlezien, 2019). Yet, as Althaus et al. (2011, p. 1065) criticize, "[a] wide range of political science scholarship still proceeds as if mass publics responded to 'real world' events and conditions. Few individuals have a direct view of events, and so most people respond instead to mediated realities constructed for them by news outlets." Therefore, it is problematic to treat proportions of fulfilled manifesto promises as the primary indicator that voters use to assess government performance retrospectively. We need a better understanding of the type of information that voters receive regularly from the news. If media mainly focus on unfulfilled or ongoing/announced, but not (yet) kept promises, the coverage could reinforce the common opinion that parties break campaign pledges, even though most existing studies report high degrees of pledge fulfillment.

Theory and Hypotheses

Having summarized the crucial role of media in the mandate vision of democracy, I next outline expectations and formulate testable hypotheses. The argument relates to the frequency of reports throughout the electoral cycle (i.e., the time between two elections), coverage of broken relative to fulfilled promises, and changes over time.

First, it is essential to clarify the definition of election pledges and distinguish between different types of statements on political promises. This study employs a broader conceptualization of campaign promises than scholars working on manifesto pledge fulfillment (Naurin et al., 2019a; Thomson et al., 2017). I define pledges as testable or subjective statements about promised actions or outcomes by political parties or politicians.[2] I use the terms "pledge" and "promise" interchangeably throughout the paper. Broadening the definition for the analysis of newspaper coverage and considering non-testable pledges seems reasonable because voters tend to have a wider definition of pledges than scholars (Dupont et al., 2019; Thomson, 2011).

Second, I separate coverage of political promises into three categories: "broken promise", "fulfilled promise", "ongoing promise". The class "fulfilled promise" includes sentences with a term indicating a promise and a word that indicates fulfillment. The class "broken promise" describes an action or outcome that points to the breaking of a promise or that politicians or parties failed to deliver on a pledge. Statements about "ongoing promises" mention a promise announced by a party or politician, provide updates about pledges made before an election, or broadly comment on a pledge. Yet, sentences about ongoing promises usually do *not* include details on the fulfillment or breaking of a pledge. For instance, the Irish Times reported that the Irish Prime Minister Leo Varadkar "pledges to increase threshold for top tax rate to 50,000 EUR" (Irish Times, 18 November 2018). This statement contains information on a promise, but readers do not receive information on its breaking or fulfillment. Ongoing promises can be important for the public perception of keeping and breaking pledges. If news outlets publish many reports on political promises, but fewer articles on broken or fulfilled pledges, readers could receive the impression that politicians "promise ever more" (Håkansson & Naurin, 2016), but fail to fulfill these policy proposals.

Turning to the hypotheses, I expect temporal variation and a dynamic pattern in the amount of pledge coverage throughout the electoral cycle. I define the electoral cycle as the period between two general elections. After *Election t*, media coverage on promises should be low (*Phase 1*). In single-party cabinets, important personnel decisions must be made before policies can be enacted. 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 multiparty systems, before entering a coalition, parties negotiate with each other and regularly draft coalition agreements (Müller & 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 (Ecker & Meyer, 2020). In these cases, I expect that pledge coverage remains on a low level until the government has been formed. Afterwards, parties should start fulfilling or breaking promises (*Phase 2*). The only study on the timing of pledge fulfillment (Duval & Pétry, 2019) provides anecdotal evidence that parties fulfill the largest share of pledges in the first year of an electoral cycle. As a result, coverage of promises should reach its first relative peak approximately at the first quarter of the cycle. Reports 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 could increase again. Yet, the level of reporting on pledges is likely to be lower than before "first-order" national elections.

The largest proportion of coverage is expected to occur in *Phase 4*, the time before *Election t+1*. The launch of party manifestos in the weeks or months before an election serves the purpose of making the party's positions public. News outlets compare parties' positions on salient issues, and summarize the promises by the major parties for the upcoming legislative period (Kostadinova, 2017). Additionally, politicians outline their plans for the future in interviews or press releases before elections, and voters become more interested in parties' pledges as an election draws closer (Bischof & Senninger, 2018). Besides the increase in coverage of promises for the upcoming legislative period, politicians and journalists also retrospectively judge the fulfillment or breaking of pledges in the current cycle. Taking all of these factors together, coverage of ongoing, broken, and fulfilled promises should increase exponentially before general elections.

H1 (Electoral Cycle Hypothesis) *Media coverage about promises increases until approximately the end of the first quarter and peaks toward the end of an electoral cycle.*

The second expectation relates to the coverage of broken and fulfilled promises. If there was congruence between the empirical evidence (e.g., Naurin et al., 2019a; Thomson et al., 2017) and news coverage, we would expect the media to devote similar attention to broken and fulfilled promises, or even to cover fulfilled promises more than broken promises. However, findings from psychology, economics, biology, and anthropology show consistent evidence of negativity bias (for extensive reviews see Cacioppo & Garnder, 1999; Rozin & Royzman, 2001; Soroka & McAdams, 2015). Humans react more strongly to negative than to positive information, and media coverage tends to be negative. This bias seems to result from journalistic norms of cynicism toward politics as well as consumers' preferences for negative news (Soroka et al., 2019; Trussler & Soroka, 2014). Lab experiments suggest that citizens are more likely to select news stories about election campaigns that contain negatively framed information (Meffert et al., 2006). Similarly, Soroka (2014, pp. 96–101) shows that newsstand sales of newspapers decrease when the cover uses positive rather than neutral or negative tone.

For three reasons, I posit that the media reporting of pledges tends to focus more on broken than on fulfilled promises. First, a broken promise – even if the pledge is not highly salient – could simply be more "newsworthy" than a fulfilled pledge of low importance, given that humans react stronger to negative information. Second, journalists usually aim to report critically on policy processes and outcomes. The "adversary model" assumes that journalists maintain a critical relationship to parties and politics and that journalists act as watchdogs. According to this model, critical coverage protects the public from misuse of power by politicians (e.g., Blumler & Gurevitch, 1995; Eriksson & Östman, 2013). The press faces incentives to cover what elected representatives are doing wrong, which implies that instances in which politicians did their job do not receive as much media attention. Thus, journalists could have higher intrinsic motivations to report on broken promises – a means to inform the public about grievances – rather than uncritically highlighting achievements. Third, 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). Pointing to a broken promise provides the opposition with a good opportunity to criticize the incumbent and to get this short message reported in the news. Hypothesis 2 follows from these considerations:

H2 (Unfulfilled Promises Hypothesis) *News outlets report more extensively about unfulfilled promises than about fulfilled promises.*

The third hypothesis assumes that negativity in pledge reporting has increased in recent years. First of all, attitudes toward parties and politicians have changed. For example, trust in politicians and political institutions has decreased over time (Clarke et al., 2018), and many countries have seen an increase in political polarization and the rise of populist parties criticizing the political "establishment" (Mudde & Kaltwasser, 2018). These factors could have contributed to a more negative attitude toward parties' ability to deliver on their promises. The media might also have changed their style of reporting over time. "Political reporting is more imbued with qualities of challenge, vigilance, criticism, and exposure at the expense of giving credit where it may be due" (Blumler, 1997, p. 400).

Evidence from the United States shows that negativity in Presidential campaigns has increased over time. For instance, news stories increasingly cover negative advertisements. Since negative ads are far more likely to receive media coverage, candidates have more incentives to develop negative campaigns (Geer, 2012; Ridout & Smith, 2008). I posit that the coverage of election promises has evolved in similar ways. As Håkansson and Naurin (2016, p. 395) note, the "inherently moral aspect of promises […] provides a foundation for media narratives of political success and/or failure." Media outlets face financial incentives to react to readers' interests in negative news. In contrast, politicians know that they have better chances to get their voice heard in the media when they point to the failures by their competitors. For all these reasons, I expect that media coverage of broken promises has increased over time. Having collected media coverage of promises from 1979 until 2017 allows to test the following hypothesis:

H3 (Increasing Negativity Hypothesis) *The focus on broken relative to fulfilled promises has increased over time.*

Case Selection

Newspapers frequently set the agenda and provide comprehensive coverage of domestic and international political events (Harder et al., 2017; Roberts & McCombs, 1994). Pledge coverage in print outlets is very likely to be replicated on the newspapers' websites and shared via their social media accounts. Even though the influence of digital media has increased, citizens still consult newspapers for information about politics (e.g., Fieldhouse et al., 2018; Fournier et al., 2015; McAllister et al., 2016). Besides the vital agenda-setting function, 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. I opt for a comparative approach using outlets from 33 electoral cycles in Australia, Canada, Ireland, and the United Kingdom.

Although the sample is limited to English speaking parliamentary democracies, the countries offer considerable variation in institutional settings and the 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: the United Kingdom and Canada employ a first-past-the-post (FPTP) system with single-member constituencies. Australia uses the Alternative Vote, and elections in Ireland are conducted under the Single-Transferable Vote system. As a consequence of the electoral system, the effective number of parliamentary parties and the number of government parties differ across the sample.[3] The sample includes both broadsheet and tabloid newspapers. Moreover, all countries have been analyzed previously in terms of pledge fulfillment and these studies concluded that governments fulfilled more promises than commonly assumed (e.g., Carson et al., 2019; Naurin et al., 2019a; Pétry et al., 2018; Royed, 1996; Thomson & Costello, 2016). In sum, the countries vary in terms of party-system and constitutional features, the degree of federalism, and government types. Finding similar patterns of media coverage across these four institutionally different contexts would strengthen the validity and generalizability of the conclusions.

Data and Classification

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 (SI Section B). Figure 1 plots the data availability of the newspapers included in the analysis, along with the number of sentences on promises. The average newspaper availability amounts to 25.3 years (median: 27 years; standard deviation: 6 years).[4] The text corpus covers ten full cycles in Canada and Australia, seven cycles in the United Kingdom, and five electoral cycles in Ireland. The corpus includes 437,901 relevant sentences from 302,770 newspaper articles.

Analyzing almost half a million sentences mentioning promises and parties poses several challenges. A hand-coded approach is beyond the 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 relating to promises. I opt for a combination of human judgment and quantitative text analysis to filter relevant sentences from the corpus (Grimmer & Stewart, 2013; Soroka & Wlezien, 2019). Having retrieved the newspaper articles, the relevant content is extracted in several steps. Text pre-processing and the classification were conducted using the *quanteda* R package (Benoit et al., 2018). All articles and meta information on each article are imported into a text corpus. The corpus is reshaped to the level of sentences resulting in over 15.7 million sentences. Since the main interest lies in sentences that mention promises, I only selected statements that contain keywords indicating a pledge.

To classify promise-related statements, I construct a dictionary with terms indicating an election pledge as well as the fulfillment/breaking of the promise. Statements on broken and fulfilled pledges consist of a word indicating a promise and a verb in *past tense* that

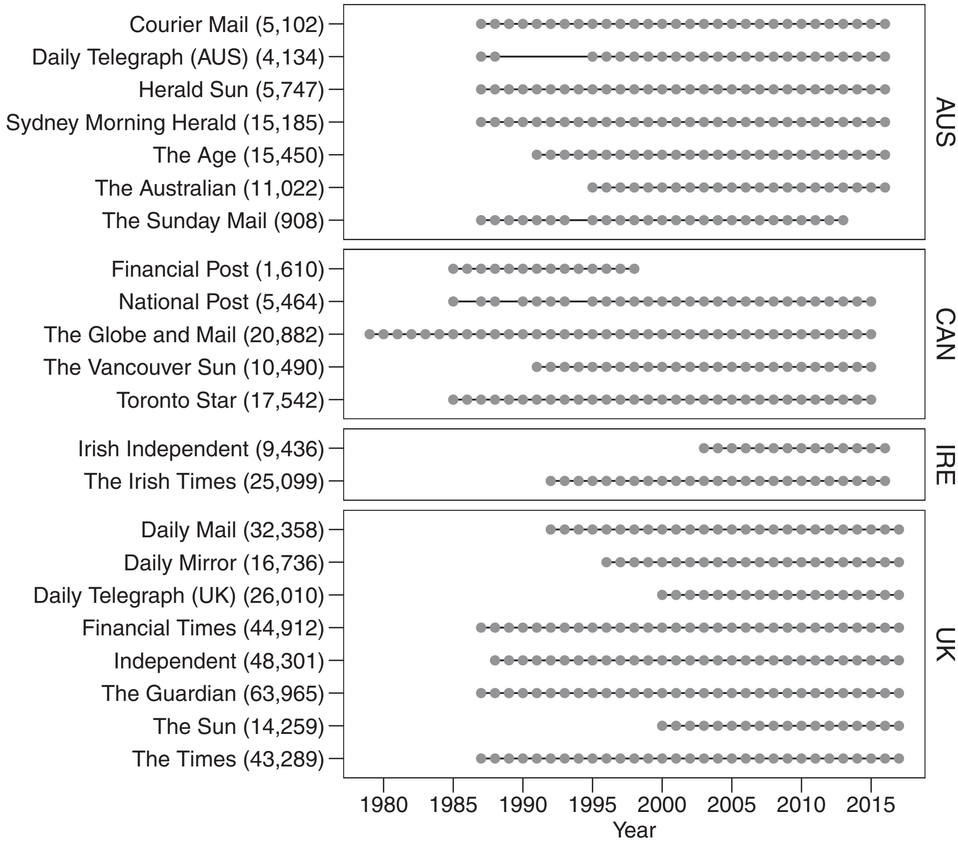


Figure 1. Newspapers included in the analysis (number of promise-related sentences in parentheses).

emphasizes either the breaking or fulfillment of a promise. To separate "ongoing promises" from sentences on nonpolitical promises, I apply a part-of-speech tagger to all available sentences from party manifestos in the United Kingdom between 1945 and 2017 and extract the 75 most frequent verbs that appear in sentences mentioning a political promise. Sentences that contain a promise-related term and one of these verbs, or mention a domestic party or Member of Parliament (MP) are classified as "ongoing promises". Selecting verbs and named entities (i.e., names of MPs and parties) as a filter for ongoing promises helps to removes sentences that mention promises which might not relate to politics. To sum up, the classes "broken" and "fulfilled" include terms describing the fulfillment or breaking of a pledge, whereas "ongoing promises" do not exclude such terms. "Ongoing promises" usually describe an action or outcome that has been announced or is in the process of being undertaken by a political party. Yet, some of these sentences may – to different degrees – be informative with regard to whether the party is or has been able to deliver on a promise.

For two reasons, I selected only a few keywords (see Table A3). First, the classification limited to key terms ensures that sentences labeled as broken or fulfilled relate to the concept under investigation. A more extensive collection of keywords increases the proportion of false-positive classifications. Recent research shows that a small dictionary can outperform larger collections of keywords (Muddiman et al., 2019). Second, sentences that summarize a promised policy action or outcome and contain information on the fulfillment or breaking provide readers with comprehensible information about parties' ability to deliver on their promises. Given that many citizens lack accurate knowledge about specific pledges made by political parties (SI Section A), it seems to be appropriate to focus on sentences that state explicitly that a political actor has promised a policy action or outcome. To be clear, even though fulfilled or broken promises can be expressed differently, this keyword-based approach allows for a transparent and intuitive classification.

Validation

Validation is essential when working with text-as-data approaches (Grimmer & Stewart, 2013). I assess the selection and classification of articles in four ways (SI Section C). First, the dictionary-based analysis resulted in a more reliable separation of classes than supervised classifiers when comparing the results to 400 English crowd-coded sentences (Benoit et al., 2016), randomly sampled from the text corpus (SI Section C.1).

Second, I investigate the immediate context of terms used to classify broken and fulfilled promises analyzing word co-occurrences for sentences on promises in each class (SI Section C.2). This textual network analysis offers strong evidence that the classification into the three categories indeed picks up sentences on politics, political promises, and policy-making, indicated by terms such as *budget*, *coalition*, *government*, *manifesto*, *minister*, and *policy*. Moreover, we observe clear differences in word co-occurrences between the classes "broken" and "fulfilled".

Third, I test whether the articles retrieved through the keyword search correspond to the articles about policy areas in which a promise has been made (SI Section C.3). 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 analyzed in Thomson and Brandenburg (2019). Around 50% of these articles also include words starting with *pledge or promise*. To be clear, a large share of articles on these policy issues includes keywords used for retrieving the sample of newspaper articles. The search query based on promise-related seed words appears justifiable. Articles on promises that have been classified as broken in Thomson and Brandenburg (2019) should also be more contain more terms from the keywords used to classify the breaking of promises. Indeed, the focus on the breaking of a promise is highest for the two unfulfilled pledges.

Finally, I examine whether the class of "ongoing promises" focuses mostly on political promises or whether these sentences may in fact clarify that a promise has been fulfilled (or broken) using words not included in the set of keywords for the classification. I randomly selected 400 sentences from the class of "ongoing promises" (100 sentences per country) and check whether a sentences is related to a political promise and whether a statement mentions the breaking or fulfillment. Around 80% of the random sample of "ongoing promises" indeed relates to political promises. 7% of the sentences contain information on the breaking of a promise, 3% of the statements point to the (potential) fulfillment of a pledge. Yet, in almost all of these instances, the details on the fulfillment or breaking are very vague (SI Section C.4). This finding mirrors the results of the crowd-sourced coding task. The dictionary-approach performs well in separating promise-related content into the three categories, and only very few sentences in the class of "ongoing promises' mention the fulfillment of a pledge.

Variables and Methods

In this section, I describe the dependent variables, independent and control variables, and the methodological approach. In order to test how coverage varies throughout the cycle (Hypothesis 1), I count the number of *Articles on promise-related sentences* per day for each of the three classes (ongoing, fulfilled, and broken). The dataset thus consists of an equal number of observations for "broken", "fulfilled", and "ongoing" promises. Each observation indicates how many sentences about each class have been published on a given day in a country. If no pledge-related sentence in one of the classes was published on a given day in a country, the count of sentences is set to 0. To test Hypotheses 2 and 3, I use the *Ratio of broken to fulfilled promises* as the dependent variable. 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 an equal number of sentences classified as "fulfilled" and as "broken". A value of 2 indicates that the number of sentences about broken promises was twice as large as the number of sentences about fulfilled promises.[5] In the Supporting Information I also measure the dependent variable as the logged ratio of statements on broken to fulfilled promises to avoid skewness and limit the influence of outliers (Lowe et al., 2011; Proksch et al., 2019). The conclusions remain unchanged.

I add a set of independent and control variables to the regression models. *Electoral cycle* serves as the independent variable for Hypothesis 1. It standardizes 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 because of different maximum term lengths or early elections (Müller & Louwerse, 2018). I add the second- and third-order polynomials of *Electoral cycle* to allow for curvilinear effects over time, and also use generalized additive models (GAM) to test whether results depend on the model specification or the inclusion of control variables. I add the continuous variable *Year* to all models to control for changes over time, and control for the *Government type* (single-party government or multiparty cabinet). Government approval and the economic situation could influence reports on political promises. Therefore, I also control for the logged quarterly *Change in GDP growth* (lagged by one quarter) and the average *Change in government approval* of all cabinet parties in a given quarter compared to the previous election using data provided in Jennings and Wlezien (2016, 2018). I add either dummies or random effects (depending on the regression model) for each *Country, Cycle*, and *Newspaper*. In order to make regression results interpretable in substantive terms, I report predicted/fitted counts (Hypothesis 1) and predicted/fitted values (Hypothesis 2 and 3) using the *effects* R package (Fox, 2003; Fox & Weisberg, 2019), holding the remaining covariates fixed at their mean (continuous variables) or modal value (factor variables) values.

Results

The results are presented in three steps. First, I provide evidence regarding the coverage of different types of reports on promises across the four countries throughout the electoral cycle (Hypothesis 1). Second, I test Hypotheses 2 and 3. Third, I summarize a series of robustness tests to assess the validity of the findings.

Coverage Throughout the Electoral Cycle

Does coverage of promises vary throughout the electoral cycle? Figure 2 plots generalized additive model (GAM) smoothers for the development of pledge coverage throughout the cycle for each country and class. In all countries, we observe a sharp increase in coverage before an election (i.e., when *Electoral cycle* approaches a value of 1). This finding holds for coverage on broken, fulfilled, and ongoing promises. Newspapers indeed inform voters about promises when an election approaches.

While the smoothed lines allow for temporal variation, they do not consider any control variables. The sample consists of multiparty cabinets and single-party governments, and we might expect different developments for cabinets comprised of more than one government party. More to the point, the focus on broken promises

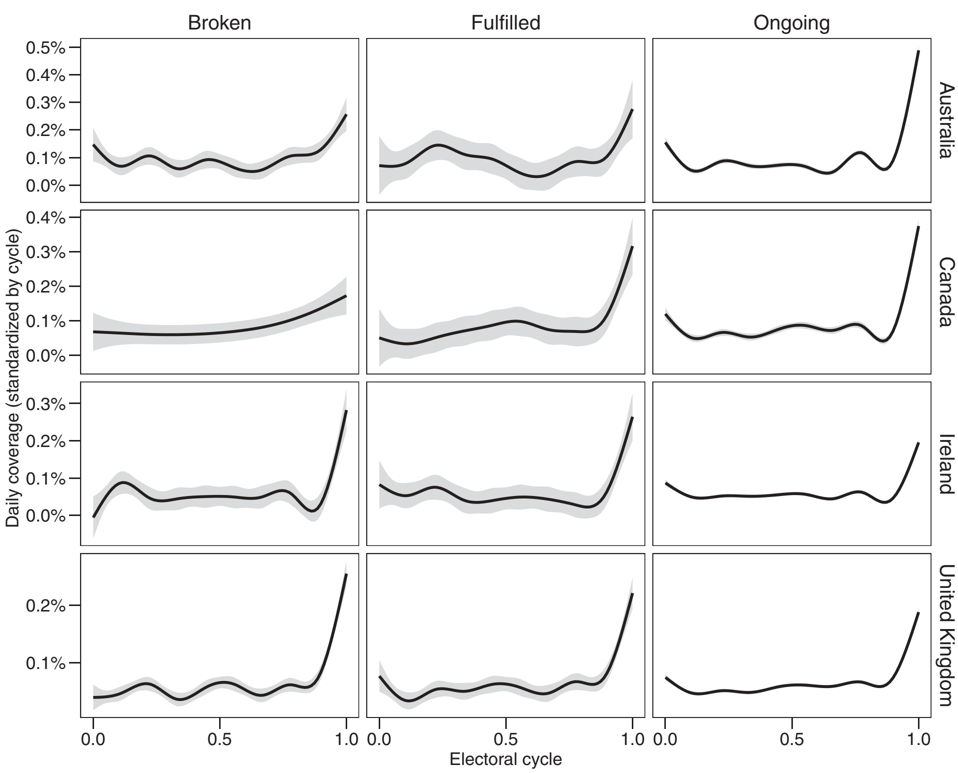


Figure 2. The daily proportion of pledge-related statements throughout the electoral cycle. The lines are based on generalized additive model (GAM) smoothers. The dates are standardized from 0 to 1 for each cycle, where 0 marks the date of *Election t*, and the date of *Election t+1* is set to 1.

in the first half of a cycle could be stronger for multiparty governments. Coalition parties need to make compromises which reduces the parties' ability to fulfill their own pledges (Naurin et al., 2019a; Thomson et al., 2017). Additionally, Duval and Pétry (2019) provide evidence from Canada that parties fulfill most pledges in the first year of the electoral cycle, which should increase coverage of fulfilled promises under single- and multiparty government.

The predicted counts of coverage conditional on the government type shows once more that coverage of all three types of political promises increases exponentially before *Election t+1* (Figure 3). The top-left panel of Figure 3 (broken promises under multiparty government) indeed suggests a small increase in the count of articles per day in the first quarter of a cycle. However, this finding should be taken with a grain of salt. The increase is rather moderate and only evident for reports on broken promises under multiparty government.[6] The plots and regressions offer partial support for Hypothesis 1: while coverage peaks before elections, we do not observe a consistently higher degree of coverage at around the first quarter of a cycle.

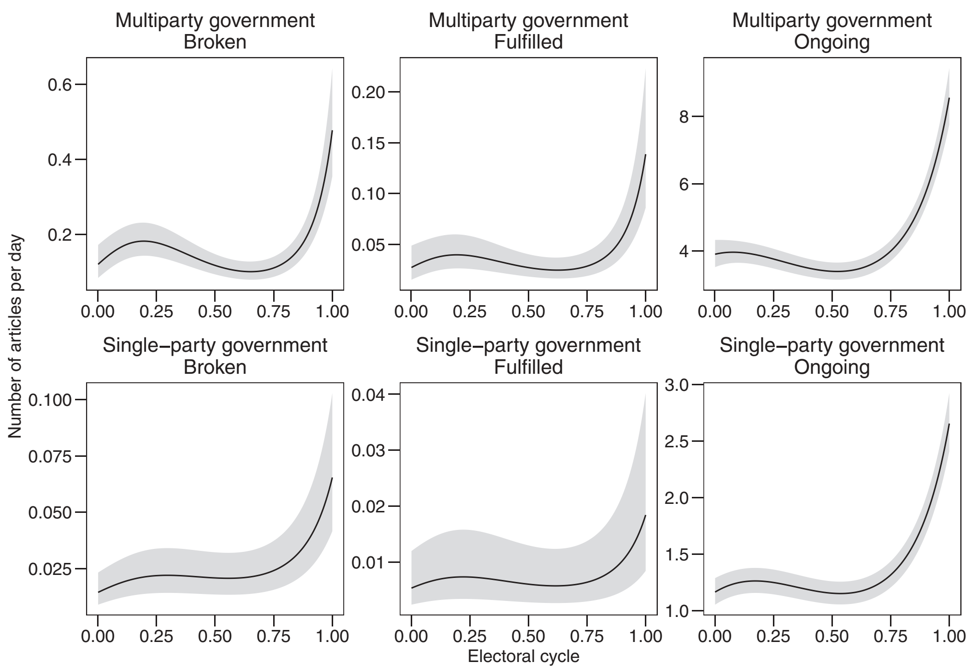


Figure 3. The predicted number of articles per day for single-party and multiparty governments, separately for each class.

The estimates are based on negative binomial regression models with the number of daily articles as the dependent variable (see Table A8). All models include first, second, and third-order polynomials for *Electoral cycle*, the lagged dependent variable, and control for changes in the economic situation, changes in government approval, and cycles. Note that the y-axes differ for each facet to show the developments within each class. Grey areas show 95% confidence intervals.

Table 1. Number of relevant sentences for each category and country (inductive dictionary, full sample).

Country	Ongoing	Broken	Fulfilled	Total
Australia	52,428 (91.2%)	3,663 (6.4%)	1,390 (2.4%)	57,481
Canada	51,755 (92.5%)	2,625 (4.7%)	1,578 (2.8%)	55,958
Ireland	31,794 (92.1%)	1,591 (4.6%)	1,127 (3.3%)	34,512
United Kingdom	273,312 (94.3%)	10,598 (3.7%)	5,803 (2.0%)	289,713

Coverage of Different Types of Promise-Related Statements

Do newspapers exert a negativity bias when reporting on political promises? Table 1 lists the number of relevant sentences for each country and category, along with the percentages of each class. Between 2.0% and 3.3% of the statements include a term indicating a fulfilled promise. The proportion of coverage on broken promises is substantively higher in all four countries, ranging between 3.7% and 6.4%. Over 90% of statements are classified as promises that mention a promise-related term and a verb co-occurring with future-related language in manifestos.

Moving beyond aggregated proportions of each class, I run multilevel linear regressions with the ratio of statements about fulfilled to broken promises per quarter in each newspaper and quarter as the dependent variable (Table 2). Model 1 uses all quarter-newspaper observations with at least one statement on broken and at least one statement on fulfilled promises.[7] First of all, tabloid newspapers exert a significantly and substantively higher negativity bias than broadsheets. Changes in the economic situation or public support do not indicate any statistical or substantive significance. The positive and

Table 2. Predicting the ratio of broken to fulfilled promises per quarter and newspaper.

	M1: Full sample	M2: Actor mentioned	M3: Logged DV
Newspaper: Tabloid	0.80**	0.63*	0.17**
	(0.28)	(0.32)	(0.06)
Canada	−0.46	−0.44	−0.17
	(0.45)	(0.58)	(0.13)
Ireland	−0.61	−0.04	−0.14
	(0.59)	(0.73)	(0.17)
United Kingdom	−0.18	−0.01	−0.01
	(0.40)	(0.52)	(0.12)
Lagged GDP change (log)	0.04	−0.12	0.02
	(0.14)	(0.17)	(0.04)
Poll change to previous election	−0.02	−0.01	−0.00
	(0.01)	(0.01)	(0.00)
Year	0.03**	0.06***	0.02***
	(0.01)	(0.02)	(0.00)
Gov. type: Single-party government	0.19	0.57	0.08
	(0.31)	(0.45)	(0.12)
AIC	7288.66	6689.23	3362.98
BIC	7353.33	6752.28	3427.65
Log likelihood	−3632.33	−3332.61	−1669.49
N	1619	1414	1619
N (Cycles)	32	32	32
N (Newspapers)	22	22	22

***p < .001, **p < .01, *p < .05.

Model 1 uses the full sample of quarters with at least one sentence on broken and one sentence on fulfilled promises. Model 2 considers only sentences that also mention a political party or MP serving in the respective cycle. Model 3 uses a logged dependent variable of the ratio. Models include random intercepts for each newspaper and cycle. Intercepts omitted from table. Standard errors in parentheses.

statistically significant coefficient for *Year* suggests an increase in the focus on broken promises over time. Results remain very similar when limiting the sample to sentences that directly refer to a political party or a Member of Parliament (Model 2) or using the lagged ratio of broken to fulfilled promises (Model 3 of Table 2; see also SI Section D).

Figure 4 shows the values of the ratio of broken to fulfilled promises based on the coefficients from Model 1. Values exceeding 1 indicate a higher emphasis on broken promises. On average, newspapers across the four countries cover broken promises 2 to 2.5 times more often than fulfilled promises (Figure 4a). The result is significant for Australia, Canada, and the United Kingdom. As Table 2 and the predicted values for the type of newspaper indicate (Figure A14), broadsheets exert a smaller negativity bias than tabloids. The fact that only Irish broadsheet papers were available for a sufficiently long period might explain the slightly lower point estimate and larger standard errors. When controlling for outliers using the logged ratio, the effect is also statistically significant for the Irish outlets. These results and alternative model specifications offer strong support for Hypothesis 2. Newspapers focus much more on broken than on fulfilled promises.

As posited in Hypothesis 3, statements on broken promises, relative to fulfilled promises, should have increased throughout time. Figure 4b plots the predicted ratio of broken to fulfilled promises over time. Whereas coverage seemed to be more balanced in the 1980s, the focus on broken, relative to fulfilled promises has increased from around 1.5 times higher coverage of broken promises to a 2.5 times higher coverage in the 2000s. Figure A15 repeats the analysis with loess regressions and points to similar conclusions. With some fluctuations, media outlets tend to focus more on broken promises since the 2000s. Importantly, the fulfillment of campaign pledges has not decreased in recent years, as a reanalysis of Thomson et al.'s (2017) comparative study underscores (Figure A16). Thus, the media has not increased the focus on broken promises as a reaction to a decline in pledge fulfillment. Figure A17 shows that most articles were published in broadsheet

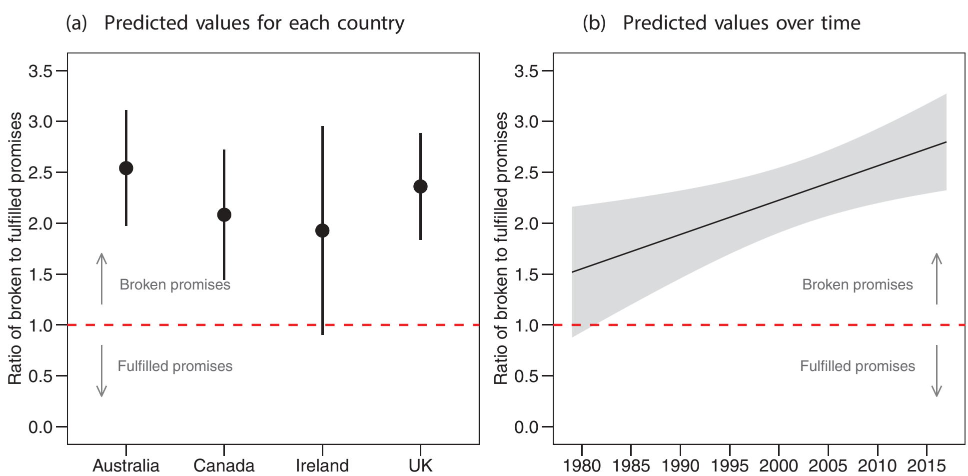


Figure 4. The predicted ratio of reports on broken to fulfilled promises per quarter and newspaper. The estimates are based on Model 1 of Table 2. A ratio of 1 implies that the number of statements on broken promises equals the number of statements on fulfilled promises. Vertical bars and shaded areas show 95% confidence intervals.

newspapers and that the share of statements from broadsheets relative to the more negative tabloids has even increased. The higher focus on broken promises in recent years is not caused by newspaper selection bias (which would imply a decline in articles from broadsheets relative to tabloids).

Turning to differences across newspapers, Figure 5 shows bootstrapped ratios of reports on broken promises to reports on fulfilled promises for each newspaper. The ratios exceed the value of 1 (which would suggest equal coverage of broken and fulfilled promises) across all 22 papers. All outlets tend to exert a negativity bias. Yet, we observe differences between tabloids and broadsheets. As expected, the focus on broken promises is much stronger in tabloid newspapers, especially in the United Kingdom, a country with high newspaper polarization (Hallin & Mancini, 2004; Ladd & Lenz, 2009; Wring & Deacon, 2018). For instance, The Sun and the Daily Mirror cover broken promises between 3.5 and 4.5 more than fulfilled promises.

Robustness Tests

I conducted various robustness checks related to the classification, measurement, or modeling choice. First, using a dictionary-based classification offers advantages (straightforward interpretation and transparency), but also disadvantages (the choice of keywords can influence results). One way of dealing with this disadvantage is the construction of more than one dictionary. I developed an alternative dictionary derived from a comparison with human coding of sentences on political promises. The terms that maximize correspondence between the human coding and dictionary classification contain slightly different keywords and do not limit the terms for broken and fulfilled promises to

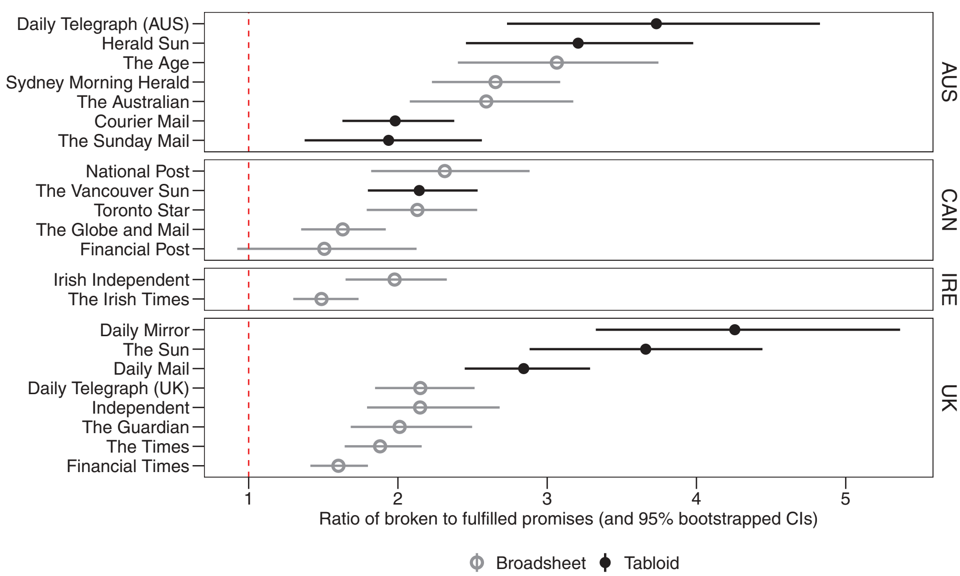


Figure 5. The quarterly focus on broken promises relative to fulfilled promises in newspapers. The points show the mean values, the horizontal bars depict 95% bootstrapped confidence intervals.

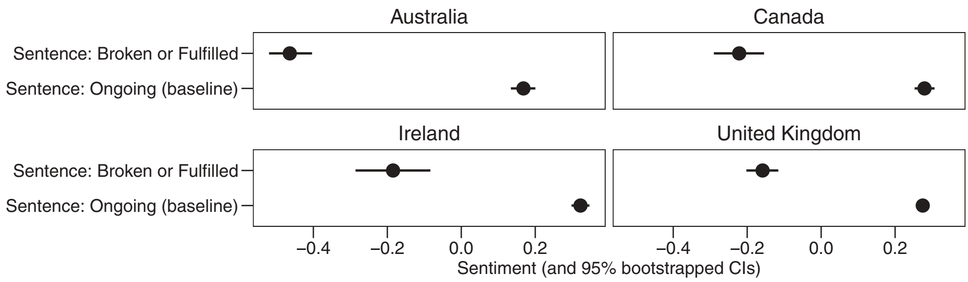


Figure 6. Aggregated sentiment (per quarter and newspaper) for sentences classified either as "broken" or "fulfilled", against the baseline of sentences classified as "ongoing".
The points show the mean values, the horizontal bars depict 95% bootstrapped confidence intervals. Lower values suggest more negative sentiment.

verbs in past tense. Despite the different choice of words, the aggregated proportions correlate highly and regression results remain the same (SI Section E.1).

Second, I measure negativity bias by conducting a sentiment analysis (for a similar procedure, see Duval, 2019). If the media focus on negative aspects when reporting on promises, we would expect the tone of sentences classified *either* as "broken" *or* "fulfilled" to be more negative than the tone of sentences classified as "ongoing" promises. I apply the Lexicoder Sentiment Dictionary (Young & Soroka, 2012), a frequently used sentiment dictionary for political and economic news, to all sentences in the corpus.[8] If the media did *not* exert a negativity bias, we would expect sentences on broken *or* fulfilled promises not to differ substantively from sentences classified as ongoing promises. In all four countries, however, the average scores for sentences on broken promises *or* fulfilled promises are significantly and substantively less positive than the baseline category (Figure 6).

This finding also holds when running multilevel regressions with the sentiment of each sentence as the dependent variable and after adding relevant control variables. The estimated difference between the baseline of "ongoing promises" and sentences on "broken" or "fulfilled" promises corresponds to 60% of the standard deviation of the sentiment variable, indicating a substantive negativity bias. Moreover, we observe a development toward higher negativity in sentences on broken and fulfilled promises in recent years, mirroring the finding that the focus on broken promises relative to fulfilled promise has increased (SI Section E.2).

Finally, investigating the most "unique" terms for sentences classified as "broken" and "fulfilled" reveals that many words in the class of "fulfilled promises" (e.g., *never*, *little*, *small*) have a negative connotation and might instead refer to the breaking of a pledge (SI Section E.3). Given that the dictionary approach tends to overestimate the proportion of fulfilled promises (SI Section C.1), the gap between reports on broken and fulfilled promises reported in this paper should be interpreted as conservative estimates of the negativity bias.

Discussion

Do the media exert a negativity bias when reporting on campaign promises? This paper constitutes the most comprehensive study on media coverage of election pledges to date.

Results allow for three conclusions. First, reporting on pledges reaches its maximum level of coverage before general elections. This finding is reassuring for the underlying assumptions of promissory representation and political accountability (Mansbridge, 2003; Powell, 2000): voters indeed receive information on broken and fulfilled pledges as well as promises for the upcoming legislative cycle. Second, I uncover a consistent negativity bias in media coverage of promises. News outlets usually report at least twice as much on broken promises than on fulfilled promises. All newspapers analyzed in this paper exert a negativity bias, with some of the tabloids focusing 3.5 to 4 times more on broken pledges. Alternative specifications of broken and fulfilled pledges and a sentiment analysis offer additional support for this finding. Third, the strength of the negativity bias seems to have increased over time.

Although the selected countries show variation in terms of the party system, federalism, and government types, the study is limited to a selection of English speaking parliamentary democracies. Future research should analyze more countries that have experienced multiparty or minority governments in order to assess whether the countries are indeed generalizable. 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.

Coverage in newspapers could be compared to the emphasis placed on issues in party manifestos to investigate whether salient pledges in manifestos also receive more media attention before an election. If the differences between salience in manifestos and newspapers are large, we might have even more doubts as to whether the proportion of fulfilled manifesto pledges changes voters' perceptions of government performance. Future research should also combine newspaper coverage of promises with the theory of issue ownership (Petrocik, 1996). If a party is perceived as being competent on a particular issue, the media might report more often and more negatively if the party fails to deliver on a pledge in this policy area.

The results have important implications for the linkage between parties and voters. Media outlets tend to stress political failures more than political success when covering promises. The extensive coverage of ongoing promises might – at least partially – explain why so many voters do not believe that politicians keep their promises (ISSP Research Group, 2018). Since the largest share of coverage is not devoted to broken or fulfilled pledges, voters could have the feeling that parties "promise ever more" (Håkansson & Naurin, 2016) without knowing about the actual performance of parties in office. Besides, the finding that most statements on promises do not contain information on their breaking or fulfillment might help us to understand why citizens' evaluations of pledges often tends to be inaccurate (Duval & Pétry, 2018; Thomson & Brandenburg, 2019). I hope that the findings of this paper encourage researchers to pay more attention to the media when analyzing the mandate-to-policy linkage.

Notes

1. A re-analysis of these studies reveals that, on average, 57% of respondents made an inaccurate evaluation or did not know whether a pledge has been fulfilled (SI Section A). Evidence from a candidate study suggests that even politicians vastly underestimate the number of pledges in their own party's manifesto (Naurin, 2016).
2. Thomson et al. (2017, p. 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."

3. Note that the cabinets between the Liberal Party of Australia and the National Party of Australia can be regarded either as coalitions or as single-party governments. Even though both parties are separate organizations, the Liberals and Nationals maintain informal cooperation and formal agreements. I run the analyses for the type of government two times: once by treating both the Liberals and Nationals as only one party and once by coding the cabinets consisting of the two parties as a coalition. The results do not change.
4. I merge weekday and Sunday editions of the same newspaper. Results remain the same when Sunday editions are treated as separate outlets.
5. I only consider newspaper-quarter observations with at least one statement on broken and one statement on fulfilled promises.
6. Counting the Liberal Party of Australia and the National Party of Australia as only one party results in high multicollinearity between the factor level "Ireland" and the type of government (coalition vs. single-party government) because in this case – apart from the coalition in the United Kingdom between 2010 and 2015 – the sample consists only of Irish coalition governments. However, coefficients for the central variables of interest do not change depending on the coding of Australian cabinets. Moreover, the shapes of pledge coverage throughout the electoral cycle conditional on the government type remain similar when treating cabinets consisting of the Liberals and Nationals as single- or multiparty governments (Figures A12 and A13).
7. Because not every newspaper published at least one statement on broken and one statement on fulfilled promises in a given quarter, the dataset contains some missing values. This makes a panel-data analysis unfeasible. By controlling for the year and adding random effects for each newspaper and cycle, I try to take into account time- and cycle-specific circumstances.
8. The dictionary consists of 1,709 positive and 2,858 negative terms, as well as their respective negations.

Acknowledgments

I am grateful to Ken Benoit, Constantine Boussalis, Shaun Bowler, Rory Costello, Alan Duggan, Fabio Ellger, Theresa Gessler, Fabrizio Gilardi, Heike Klüver, Liam Kneafsey, Gail McElroy, Sven-Oliver Proksch, Kohei Watanabe, as well as participants at the Friday Seminar Series at Trinity College Dublin, the 2018 Annual General Conference of the European Consortium of Political Research in Hamburg, the 2018 Annual Conference of the Political Studies Association of Ireland in Limerick, and the 2019 Annual General Conference of the European Political Science Association in Belfast for insightful feedback. I would also like to thank the editor Claes de Vreese and three anonymous reviewers for their excellent comments and improvement suggestions.

Disclosure Statement

No potential conflict of interest was reported by the author.

Funding

Stefan Müller received financial support from an Irish Research Council (IRC) Government of Ireland Postgraduate Scholarship [Grant no: GOIPG/2016/1278].

Data Availability Statement

The data described in this article are openly available within the Harvard Dataverse Network at https://doi.org/10.7910/DVN/POCRVT.

Open Scholarship

This article has earned the Center for Open Science badges for Open Data and Open Materials through Open Practices Disclosure. The data and materials are openly accessible at https://doi.org/10.7910/DVN/POCRVT.

Notes on Contributor

Stefan Müller, (Ph.D., Trinity College Dublin) is Assistant Professor and Ad Astra Fellow in the School of Politics and International Relations at University College Dublin. His research interests include political representation, party competition, political communication, public opinion, and computational social science.

ORCID

Stefan Müller http://orcid.org/0000-0002-6315-4125

References

Althaus, S. L., Swigger, N., Chernykh, S., Hendry, D. J., Wals, S. C., & Tiwald, C. (2011). Assumed transmission in political science: A call for bringing description back in. *The Journal of Politics*, *73*(4), 1065–1080. https://doi.org/10.1017/S0022381611000788

Belchior, A. M. (2019). The effects of party identification on perceptions of pledge fulfilment: Evidence from Portugal. *International Political Science Review*, *40*(5), 627–642. https://doi.org/10.1177/0192512118791298

Benoit, K., Conway, D., Lauderdale, B. E., Laver, M., & Mikhaylov, S. (2016). Crowd-sourced text analysis: Reproducible and agile production of political data. *American Political Science Review*, *110*(2), 278–295. https://doi.org/10.1017/S0003055416000058

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *The Journal of Open Source Software*, *3*(30), 774. https://doi.org/10.21105/joss.00774

Bischof, D., & Senninger, R. (2018). Simple politics for the people? Complexity in campaign messages and political knowledge. *European Journal of Political Research*, *57*(2), 473–495. https://doi.org/10.1111/1475-6765.12235

Blumler, J. G. (1997). Origins of the crisis of communication for citizenship. *Political Communication*, *14*(4), 395–404. https://doi.org/10.1080/105846097199191

Blumler, J. G., & Gurevitch, M. (1995). *The crisis of public communication*. Routledge.

Cacioppo, J. T., & Garnder, W. L. (1999). Emotion. *Annual Review of Psychology*, *50*(1), 191–214. https://doi.org/10.1146/annurev.psych.50.1.191

Carson, A., Gibbons, A., & Martin, A. (2019). Did the minority gillard government keep its promises? A study of promissory representation in Australia. *Australian Journal of Political Science*, *54*(2), 219–237. https://doi.org/10.1080/10361146.2019.1588956

Chong, D., & Druckman, J. N. (2007). Framing theory. *Annual Review of Political Science*, *10*(1), 103–126. https://doi.org/10.1146/annurev.polisci.10.072805.103054

Clarke, N., Jennings, W., Moss, J., & Stoker, G. (2018). *The good politician: Folk theories, political interaction, and the rise of anti-politics*. Cambridge University Press. https://doi.org/10.1017/9781108641357

Costello, R., & Thomson, R. (2008). Election pledges and their enactment in coalition governments: A comparative analysis of Ireland. *Journal of Elections, Public Opinion and Parties*, *18*(3), 239–256. https://doi.org/10.1080/17457280802227652

Downs, A. (1957). *An economic theory of democracy*. Harper & Row.

Dunaway, J. (2013). Media ownership and story tone in campaign news. *American Politics Research*, *41*(1), 24–53. https://doi.org/10.1177/1532673X12454564

Dupont, J. C., Bytzek, E., Steffens, M. C., & Schneider, F. M. (2019). Which kind of political campaign messages do people perceive as election pledges?. *Electoral Studies*, *57*, 121–130. https://doi.org/10.1016/j.electstud.2018.11.005

Duval, D. (2019). Ringing the alarm: The media coverage of the fulfilment of electoral pledges. *Electoral Studies*, *60*, 102041. https://doi.org/10.1016/j.electstud.2019.04.005

Duval, D., & Pétry, F. (2018). Citizens' evaluations of campaign pledge fulfillment in Canada. *Party Politics*, 1–11. Online First. https://doi.org/10.1177/1354068818789968

Duval, D., & Pétry, F. (2019). Time and the fulfillment of election pledges. *Political Studies*, *67*(1), 207–223. https://doi.org/10.1177/0032321718762882

Ecker, A., & Meyer, T. M. (2020). Coalition bargaining duration in multiparty democracies. *British Journal of Political Science*, *50*(1), 261–280. https://doi.org/10.1017/S0007123417000539

Elmelund-Præstekær, C. (2010). Beyond American negativity: Toward a general understanding of the determinants of negative campaigning. *European Political Science Review*, *2*(1), 137–156. https://doi.org/10.1017/S1755773909990269

Eriksson, G., & Östman, J. (2013). Cooperative or adversarial? Journalists' enactment of the watch-dog function in political news production. *The International Journal of Press/Politics*, *18*(3), 304–324. https://doi.org/10.1177/1940161213482493

Eshbaugh-Soha, M. (2010). The tone of local presidential news coverage. *Political Communication*, *27*(1), 121–140. https://doi.org/10.1080/10584600903502623

Fieldhouse, E., Green, J., Evans, G., Schmitt, H., Van der Eijk, C., Mellon, J., & Prosser, C. (2018). *British election study, 2017: Face-to-face survey*.

Fournier, P., Cutler, F., Soroka, S. N., & Stolle, D. (2015). *The 2015 Canadian election study*.

Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software*, *8*(15), 1–27. https://doi.org/10.18637/jss.v008.i15

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). SAGE.

Geer, J. G. (2006). *In defense of negativity: Attach ads in presidential campaigns*. The University of Chicago Press.

Geer, J. G. (2012). The news media and the rise of negativity in presidential campaigns. *PS: Political Science & Politics*, *45*(3), 422–427. https://doi.org/10.1017/S1049096512000492

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, *21*(3), 267–297. https://doi.org/10.1093/pan/mps028

Håkansson, N., & Naurin, E. (2016). Promising ever more: An empirical account of Swedish parties' pledge making during 20 years. *Party Politics*, *22*(3), 393–404. https://doi.org/10.1177/1354068814549338

Hallin, D. C., & Mancini, P. (2004). *Comparing media systems: Three models of media and politics*. Cambridge University Press.

Harder, R. A., Sevenans, J., & Van Aelst, P. (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. https://doi.org/10.1177/1940161217704969

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. https://doi.org/http://doi.10.4232/1.12994

Jennings, W., & Wlezien, C. (2016). The timeline of elections: A comparative perspective. *American Journal of Political Science*, *60*(1), 219–233. https://doi.org/10.1111/ajps.12189

Jennings, W., & Wlezien, C. (2018). Election polling errors across time and space. *Nature Human Behaviour*, *2*(4), 276–283. https://doi.org/10.1038/s41562-018-0315-6

Klüver, H., & Spoon, -J.-J. (2020). Helping or hurting? How governing as a junior coalition partner influences electoral outcomes. *The Journal of Politics*, 1–36. Online First. https://doi.org/10.1086/708239

Kostadinova, P. (2017). Party pledges in the news: Which election promises do the media report? *Party Politics*, *23*(6), 636–645. https://doi.org/10.1177/1354068815611649

Kostadinova, P. (2019). Influential news: Impact of print media reports on the fulfillment of election promises. *Political Communication*, *36*(3), 412–425. https://doi.org/10.1080/10584609.2018.1541032

Ladd, J. M., & Lenz, G. S. (2009). Exploiting a rare communication shift to document the persuasive power of the news media. *American Journal of Political Science*, *53*(2), 394–410. https://doi.org/10.1111/j.1540-5907.2009.00377.x

Lowe, W., Benoit, K., Mikhaylov, S., & Laver, M. (2011). Scaling policy preferences from coded political texts. *Legislative Studies Quarterly*, *36*(1), 125–155. https://doi.org/10.1111/j.1939-9162.2010.00006.x

Manin, B., Przeworski, A., & Stokes, S. C. (1999). Elections and representation. In A. Przeworski, S. C. Stokes, & B. Manin. (Eds.), *Democracy, accountability, and representation* (pp. 29–54). Cambridge University Press.

Mansbridge, J. (2003). Rethinking representation. *American Political Science Review*, *97*(4), 515–528. https://doi.org/10.1017/S0003055403000856

McAllister, I., Bean, C., Gibson, R., Makkai, T., & Sheppard, J. (2016). *Australian election study 2016*.

McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly*, *36*(2), 176–187. https://doi.org/10.1086/267990

Meffert, M. F., Chung, S., Joiner, A. J., Waks, L., & Garst, J. (2006). The effects of negativity and motivated information processing during a political campaign. *Journal of Communication*, *2006*(1), 27–51. https://doi.org/10.1111/j.1460-2466.2006.00003.x

Moy, P., & Pfau, M. (2000). *With malice toward all: The media and public confidence in democratic institutions*. Praeger.

Mudde, C., & Kaltwasser, C. R. (2018). Studying populism in comparative perspective: Reflections on the contemporary and future research agenda. *Comparative Political Studies*, *51*(13), 1667–1693. https://doi.org/10.1177/0010414018789490

Muddiman, A., McGregor, S. C., & Stroud, N. J. (2019). (Re)claiming our expertise: Parsing large text corpora with manually validated and organic dictionaries. *Political Communication*, *36*(2), 214–226. https://doi.org/10.1080/10584609.2018.1517843

Müller, S., & Louwerse, T. (2018). The electoral cycle effect in parliamentary democracies. *Political Science Research and Methods*, 1–8. Online First. https://doi.org/10.1017/psrm.2018.47

Müller, W. C., & Strøm, K. (2008). Coalition agreements and cabinet governance. In K. Strøm, W. C. Müller, & T. Bergman. (Eds.), *Cabinet governance: Bargaining and the cycle of democratic politics* (pp. 159–199). Oxford University Press.

Naurin, E. (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. https://doi.org/10.1111/1467-9477.12076

Naurin, E., & Oscarsson, H. E. (2017). When and why are voters correct in their evaluations of specific government performance? *Political Studies*, *65*(4), 860–876. https://doi.org/10.1177/0032321716688359

Naurin, E., Royed, T. J., & Thomson, R. (Eds). (2019a). *Party mandates and democracy: Making, breaking, and keeping election pledges in twelve countries*. University of Michigan Press.

Naurin, E., Soroka, S. N., & Markwat, N. (2019b). Asymmetric accountability: An experimental investigation of biases in evaluations of governments' election pledges. *Comparative Political Studies*, *52*(13–14), 2207–2234. https://doi.org/10.1177/0010414019830740

Patterson, T. E., & Donsbagh, W. (1996). News decisions: Journalists as partisan actors. *Political Communication*, *13*(4), 455–468. https://doi.org/10.1080/10584609.1996.9963131

Petrocik, J. R. (1996). Issue ownership in presidential elections, with a 1980 case study. *American Journal of Political Science*, *40*(3), 825–850. https://doi.org/10.2307/2111797

Pétry, F., Duval, D., & Birch, L. (2018). Do regional governments fulfill fewer election promises than national governments? The case of Quebec in comparative perspective. *Party Politics*, 1–11. Online First. https://doi.org/10.1177/1354068818787353

Powell, G. B. (2000). *Elections as instruments of democracy: Majoritarian and proportional visions*. Yale University Press.

Proksch, S.-O., Lowe, W., Wäckerle, J., & Soroka, S. N. (2019). Multilingual sentiment analysis: A new approach to measuring conflict in legislative speeches. *Legislative Studies Quarterly*, *44*(1), 97–131. https://doi.org/10.1111/lsq.12218

Reeves, A., McKee, M., & Stuckler, D. (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. https://doi.org/10.1016/j.ssresearch.2015.11.002

Ridout, T. N., & Smith, G. R. (2008). Free advertising: How the media amplify campaign messages. *Political Research Quarterly*, *61*(4), 598–608. https://doi.org/10.1177/1065912908314202

Roberts, M., & McCombs, M. E. (1994). Agenda setting and political advertising: Origins of the news agenda. *Political Communication*, *11*(3), 249–262. https://doi.org/10.1080/10584609.1994.9963030

Royed, T. 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. https://doi.org/10.1017/S0007123400007419

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*(4), 296–320. https://doi.org/10.1207/S15327957PSPR0504_2.

Shoemaker, P. J., & Vos, T. P. (2009). *Gatekeeping theory*. Routledge.

Soroka, S. N. (2006). Good news and bad news: Asymmetric responses to economic information. *The Journal of Politics*, *68*(2), 372–385. https://doi.org/10.1111/j.1468-2508.2006.00413.x

Soroka, S. N. (2012). The gatekeeping function: Distributions of information in media and the real world. *The Journal of Politics*, *74*(2), 514–528. https://doi.org/10.1017/S002238161100171X

Soroka, S. N. (2014). *Negativity in democratic politics: Causes and consequences*. Cambridge University Press.

Soroka, S. N., Fournier, P., & Nir, L. (2019). Cross-national evidence of a negativity bias in psychophysiological reactions to news. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, *116*(38), 18888–18892. https://doi.org/10.1073/pnas.1908369116

Soroka, S. N., & McAdams, S. (2015). News, politics, and negativity. *Political Communication*, *32*(1), 1–22. https://doi.org/10.1080/10584609.2014.881942

Soroka, S. N., & Wlezien, C. (2019). Tracking the coverage of public policy in mass media. *Policy Studies Journal*, *47*(2), 471–491. https://doi.org/10.1111/psj.12285

Thomson, R. (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. https://doi.org/10.1111/1475-6765.00595

Thomson, R. (2011). Citizens' evaluations of the fufillment of election pledges: Evidence from Ireland. *The Journal of Politics*, *73*(1), 187–201. https://doi.org/10.1017/S0022381610000952.

Thomson, R., & Brandenburg, H. (2019). Trust and citizens' evaluations of promise keeping by government parties. *Political Studies*, *67*(1), 249–266. https://doi.org/10.1177/0032321718764177

Thomson, R., & Costello, R. (2016). Governing together in good and bad economic times: The fulfilment of election pledges in Ireland. *Irish Political Studies*, *31*(2), 182–203. https://doi.org/10.1080/07907184.2016.1149066

Thomson, R., Royed, T. J., Naurin, E., Artés, J., Costello, R., Ennser-Jedenastik, L., Ferguson, M., Kostadinova, P., Moury, C., Pétry, F., & Praprotnik, K. (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. https://doi.org/10.1111/ajps.12313

Trussler, M., & Soroka, S. N. (2014). Consumer demand for cynical and negative news frames. *The International Journal of Press/Politics*, *19*(3), 360–379. https://doi.org/10.1177/1940161214524832

Walgrave, S., Soroka, S. N., & Nuytemans, M. (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. https://doi.org/10.1177/0010414006299098
Wlezien, C. (2017). Policy (mis)representation and the cost of ruling: U.S. presidential elections in comparative perspective. *Comparative Political Studies*, *50*(6), 711–738. https://doi.org/10.1177/0010414015626446
Wring, D., & Deacon, D. (2018). A bad press: Newspapers. In P. Cowley & D. Kavanagh. (Eds.), *The British general election of 2017* (pp. 347–384). Palgrave Macmillan.
Young, L., & Soroka, S. N. (2012). Affective news: The automated coding of sentiment in political texts. *Political Communication*, *29*(2), 205–231. https://doi.org/10.1080/10584609.2012.671234

Media Coverage of Campaign Promises Throughout the Electoral Cycle

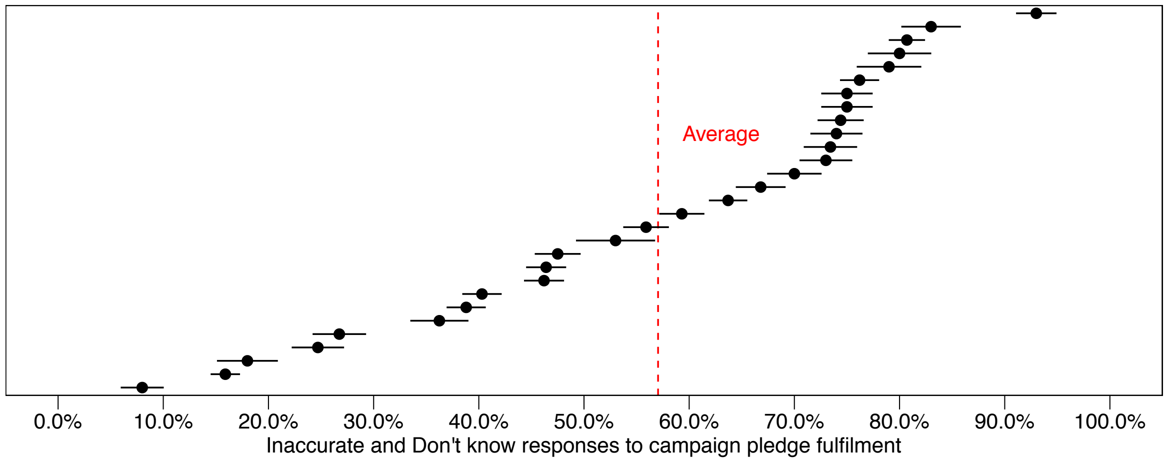
Supporting Information

Stefan Müller

A Knowledge of Pledges and Pledge Fulfillment

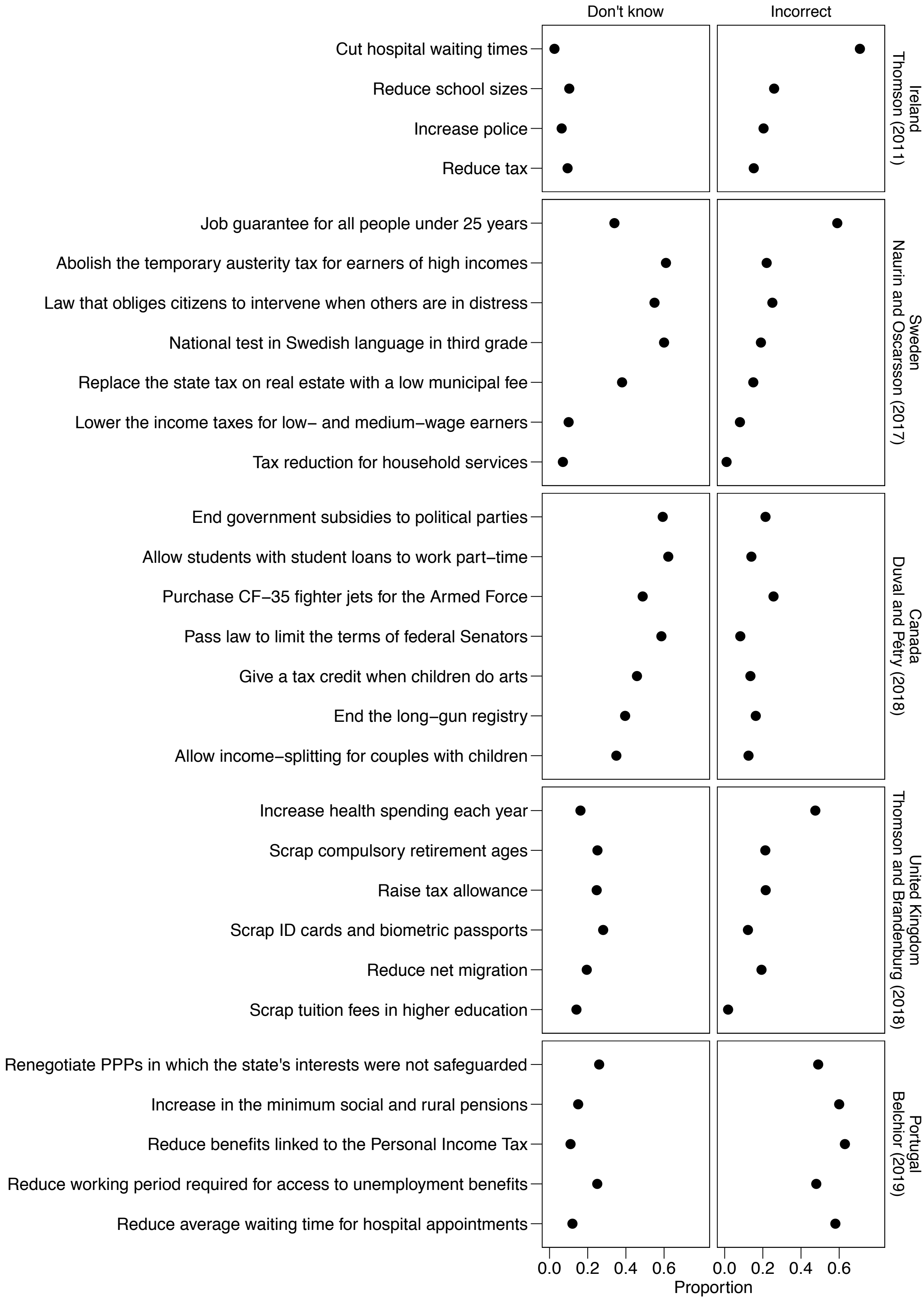
Figures A1 and A2 list the pledges included in the studies by Thomson (2011), Thomson & Brandenburg (2019), Naurin & Oscarsson (2017), Belchior (2019), and Duval & Pétry (2018), along with the proportions of respondents who evaluated the fulfillment incorrectly and the proportions of respondents who replied 'don't know'.

Figure A1: The proportions of respondents who did not know about the fulfillment 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 & Brandenburg 2019; Naurin & Oscarsson 2017; Belchior 2019; Duval & Pétry 2018).

Figure A2: The proportions of respondents who did not know about the fulfilment of a pledge and who made an incorrect evaluation, along with the exact pledge



Note: Each dot shows one pledge from the four published studies on citizens' pledge evaluations (Thomson 2011; Thomson & Brandenburg 2019; Naurin & Oscarsson 2017; Belchior 2019; Duval & Pétry 2018).

B Classifying Promise-Related Sentences

As described in the paper, I selected newspapers with the highest circulation within each country. 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 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 kept the longest version of duplicated articles in the text corpus.

Figure A3 outlines the classification and validation. 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 15.7 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 include `pledge*`, `promise*`, `guarantee*`, `assure*`, or `ensur*` which reduces the text corpus by over 97% to 512,683 sentences (for a similar classification method see, e.g., Soroka & Wlezien 2019).

Afterward, I check whether the statement (or its contextual unit of ± one sentence) mentions a party or politician. I tokenize the pledge-related sentences and 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 legislative periods not covered by *EveryPolitician*, I scraped the names and party affiliations

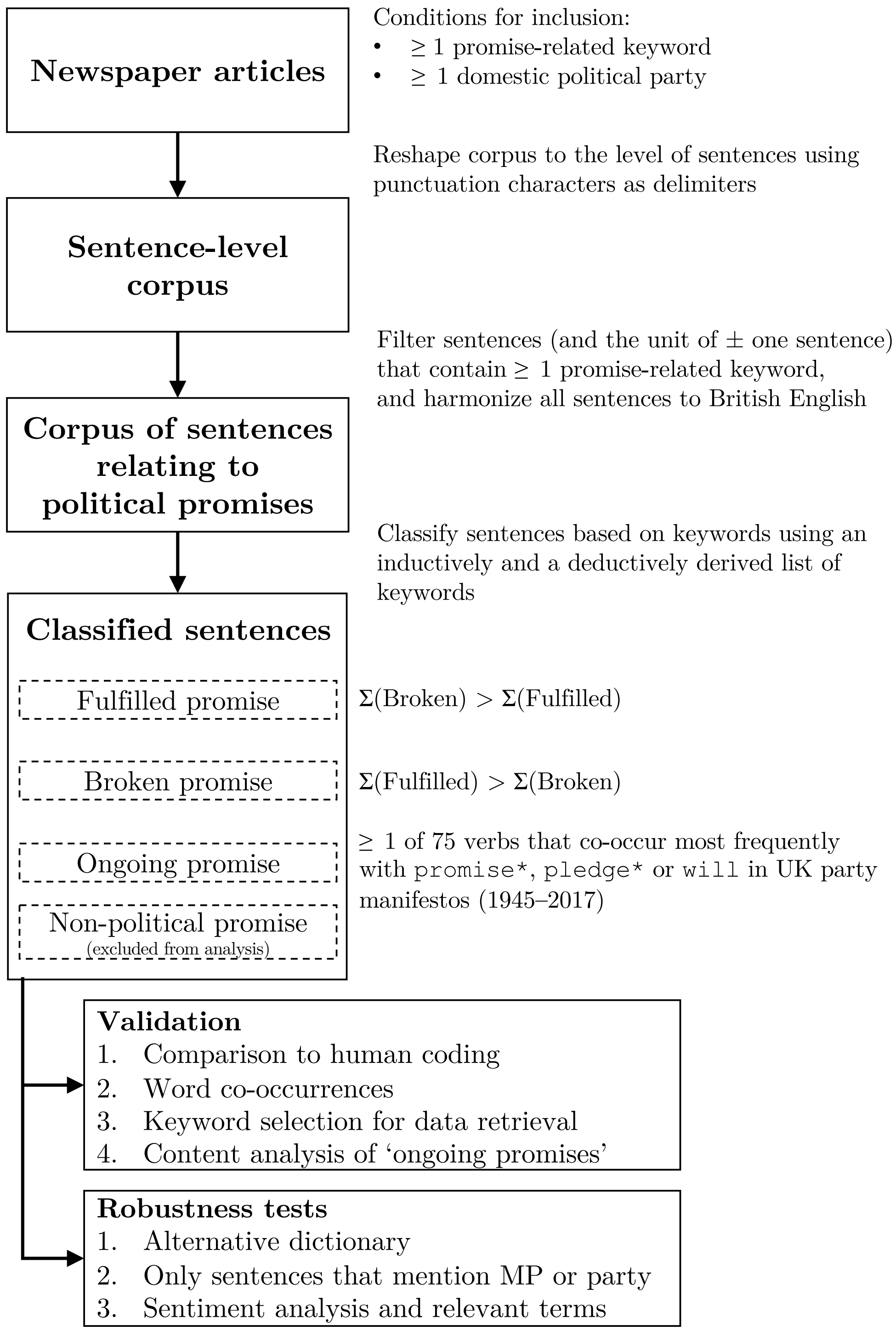
[1]For example, the following string has been used for the UK newspapers: "`promise* OR pledge* OR manifesto* AND Labour* OR Conservatives* OR Tories* OR Conservative Party* OR Liberal Democrats* OR Liberal Democratic Party* OR UKIP* OR Scottish National Party* OR Democratic Unionist Party* OR DUP* OR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

[3]https://everypolitician.org.

from the websites of the national parliaments. Note that "glob"-style wildcard matching is used to detect genitives and plural forms of parties or MPs. To minimize the number of 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 the classification only considers politicians that served in the respective electoral cycle. Next, I apply the dictionaries 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'). Sentences not containing keywords from the 'broken' or 'fulfilled' categories, but mentioning a promise-related term and a name of a party or MP or one of the 75 most frequent verbs co-occurring with terms indicating a promise in party manifestos from the United Kingdom are classified as 'ongoing'. I conduct various validation checks and robustness tests, which are outlined in detail in the paper and the following sections.

Figure A3: Retrieving, segmenting, and classifying newspaper articles

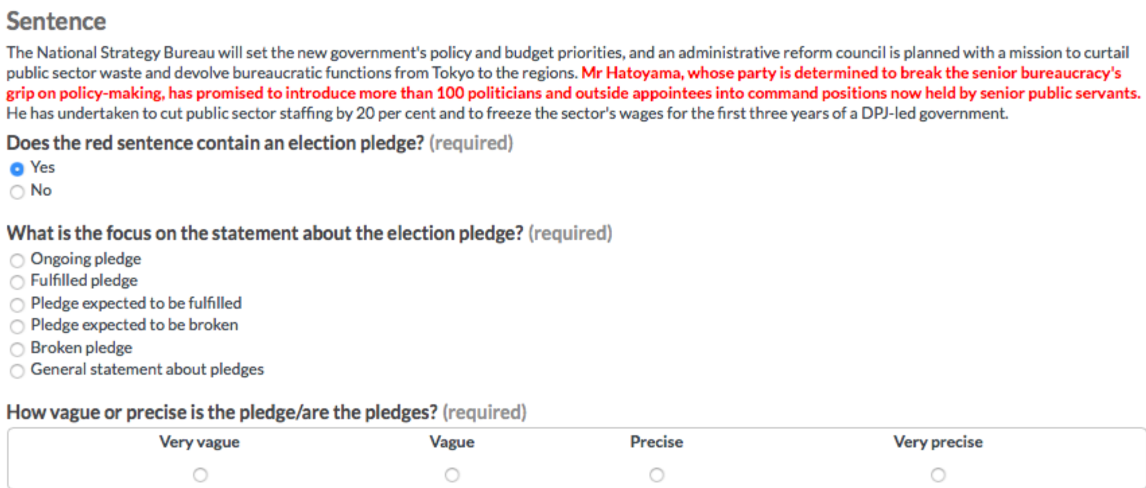


C Validation

C.1 Comparison to Human Coding

To test whether the classification into broken, fulfilled, and ongoing promises is meaningful, I compare the crowd-sourced coding (Benoit et al. 2016) of a random sample of 400 sentences to the performance of various classifiers on the same set of sentences. 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 F). 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% of all sentences are test questions with a pre-defined answer key. If respondents do not answer more than 80% of these test questions correctly, they get excluded from the job, and their previous codings are not considered in the analysis. Each sentence is coded by three workers and aggregated using the most frequent answer. Figure A4 shows a screenshot of the coding job. Workers are asked to judge the class of the sentence highlighted in red.

Figure A4: Screenshot from crowd coding job



Sentence

The National Strategy Bureau will set the new government's policy and budget priorities, and an administrative reform council is planned with a mission to curtail public sector waste and devolve bureaucratic functions from Tokyo to the regions. **Mr Hatoyama, whose party is determined to break the senior bureaucracy's grip on policy-making, has promised to introduce more than 100 politicians and outside appointees into command positions now held by senior public servants.** He has undertaken to cut public sector staffing by 20 per cent and to freeze the sector's wages for the first three years of a DPJ-led government.

Does the red sentence contain an election pledge? (required)

- Yes
- No

What is the focus on the statement about the election pledge? (required)

- Ongoing pledge
- Fulfilled pledge
- Pledge expected to be fulfilled
- Pledge expected to be broken
- Broken pledge
- General statement about pledges

How vague or precise is the pledge/are the pledges? (required)

Very vague	Vague	Precise	Very precise
○	○	○	○

Table A1 compares the dictionary-based classification with the most frequent crowd-sourced coding of the same set of sentences. I assigned the 'expected to be 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 0.64 (with a range from 0.61 to 0.69). The F1 score for the class 'fulfilled' is the lowest one, since the dictionary picked up 'false positives', which results in low precision (Table A2). The aggregated crowd codings classify 41 sentences as 'fulfilled', whereas the dictionary classifies 68 sentences as 'fulfilled'. The 'true' difference between broken and fulfilled promises is likely to increase even more if human coders had labeled all sentences. To be more precise, the estimated ratios of broken to fulfilled promises, reported in the paper and based on the dictionary approach, are likely to under- rather than overestimate the negativity bias in newspapers. The 'keyness' statistics (SI Section E.3) and the examination of the class 'ongoing promise' (SI Section C.4) provide further support for this conclusion.

Table A3 lists all words and expressions used for the inductive and deductive classification. I thank one of the anonymous reviewers for pointing out that negations might be conceptually different. For instance, 'not fulfilled' may not be the same as 'broken'. However, negations occur very rarely: fewer than 3% (broken) and 1% (fulfilled) of all matched dictionary terms within each class are negations. Negations do not influence results or the interpretation of the findings.

Table A1: Cross-table of dictionary-based classifier

	Broken	Fulfilled	Other
Crowd: Broken	146	13	52
Crowd: Fulfilled	1	33	7
Crowd: Other	63	22	121

Table A2: Performance of dictionary-based classification

Class	F1 score	Precision	Recall
Class: Broken	0.69	0.69	0.70
Class: Fulfilled	0.61	0.80	0.49
Class: Other	0.63	0.59	0.67

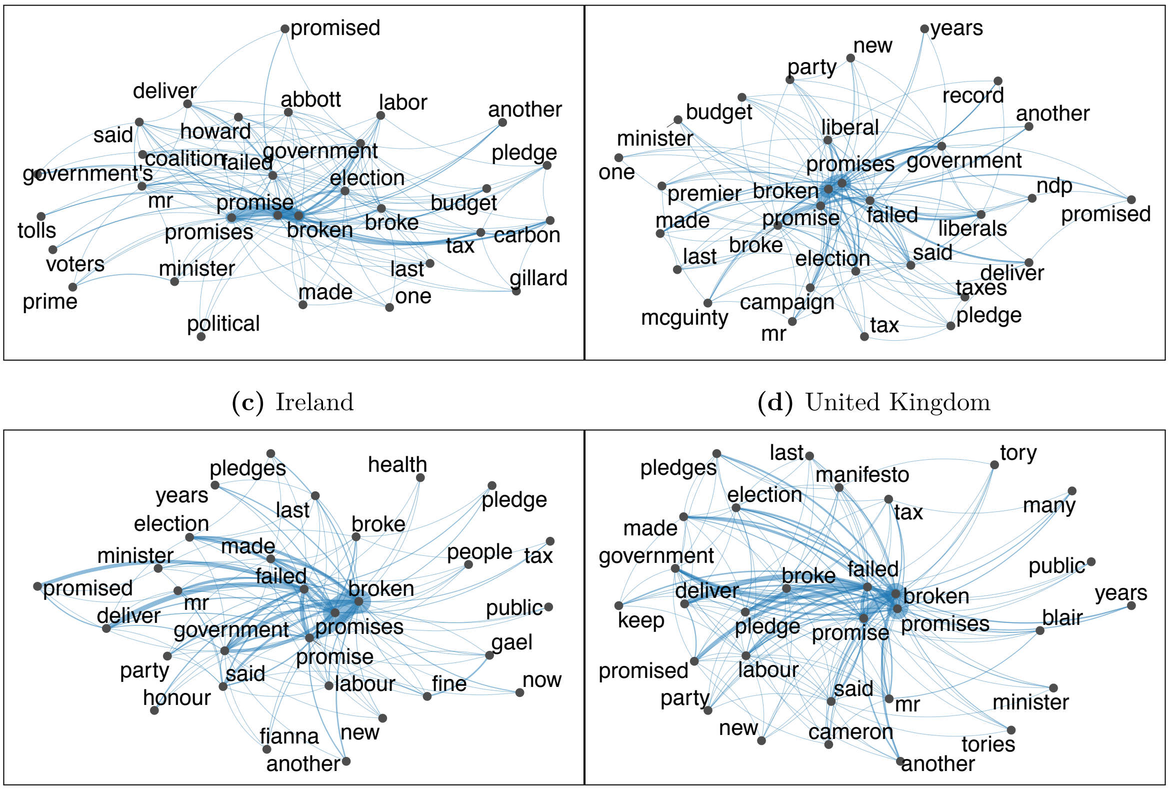
Table A3: Words used for dictionary classification

Category	Terms
Pledge	assure*; ensur*; guarantee*; pledge*; promise*
Ongoing	abolish; achieve; allow; based; become; believe; bring; build; can; continue; create; cut; deliver; develop; enable; encourage; end; ensure; ensuring; establish; existing; expand; extend; get; give; given; giving; go; help; implement; improve; include; including; increase; increased; increasing; introduce; invest; keep; made; maintain; make; making; meet; must; need; paid; pay; press; prevent; promote; protect; provide; providing; put; raise; receive; reduce; reform; remain; require; restore; review; see; seek; set; strengthen; support; tackle; take; use; want; will; work; working
Fulfilled (inductive)	delivered; fulfilled; implemented; kept; not been broken; not break; not broken; not fail*
Fulfilled (deductive)	fulfil*; kept; not break*; not broken
Broken (inductive)	broke; broken; failed; not been delivered; not been fulfilled; not been implemented; not been kept; not delivered; not fulfilled; not implemented; not kept
Broken (deductive)	broke; broken; fail* to; failure; not fulfil*; not keep*; not kept

C.2 Word Co-Occurrences

Figures A5 and A7 show network graphs of the co-occurrence of terms used to classify broken and fulfilled promises. I construct a feature co-occurrence matrix (Benoit et al. 2018) of the dictionary terms used to classify broken/fulfilled promises, retrieve the 30 most frequent features and create network plots for each country and the two classes (broken and fulfilled). Thicker lines indicate that two words co-occur (within a window of 3 features) more often. The terms clearly show that the most frequent co-occurring terms refer to policy-making, political institutions, parties, and politicians.

Figure A5: Network plots of word co-occurrences for sentences classified as 'broken promise'



(a) Australia

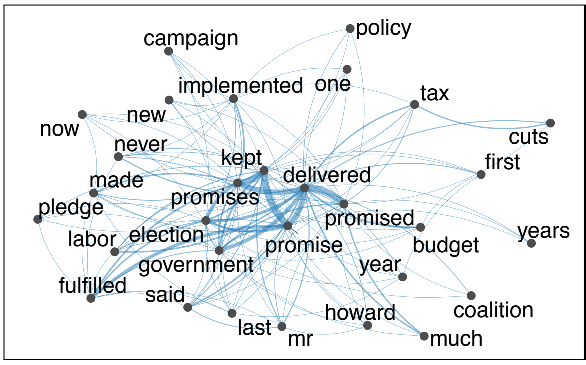
(b) Canada

(c) Ireland

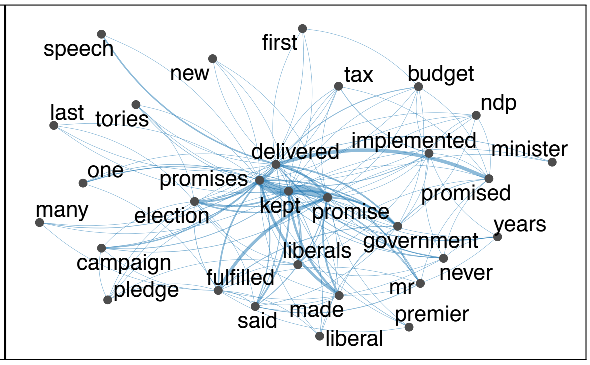
(d) United Kingdom

Figure A6: Network plots of word co-occurrences for sentences classified as 'fulfilled promise'

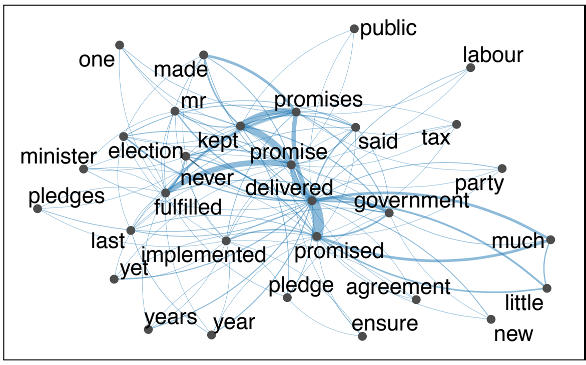
(a) Australia



(b) Canada



(c) Ireland



(d) United Kingdom

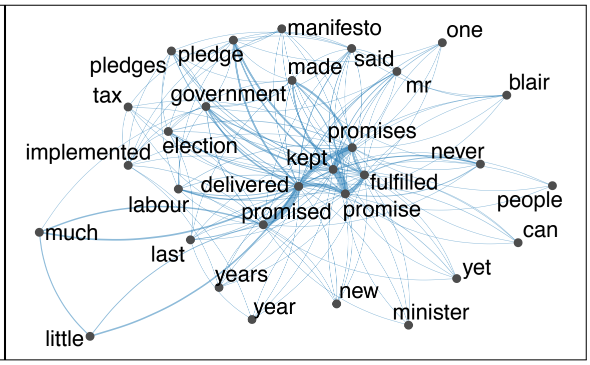
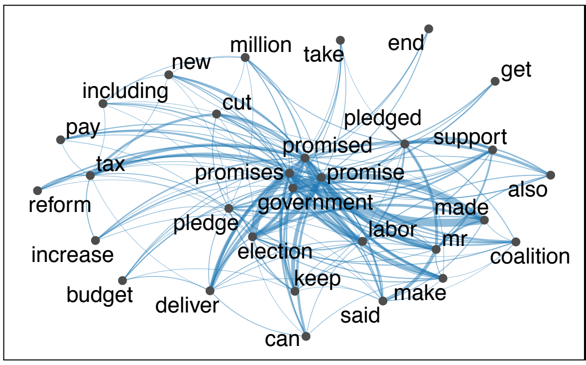
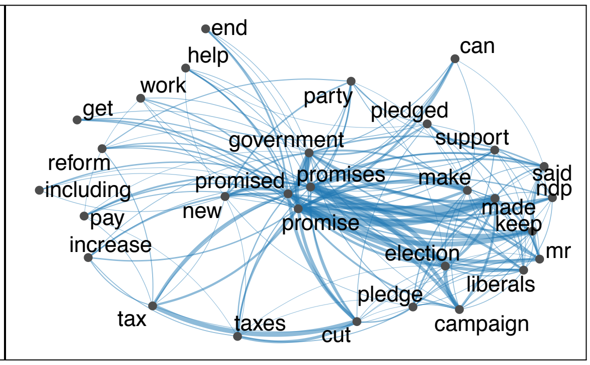


Figure A7: Network plots of word co-occurrences for sentences classified as 'ongoing promise'

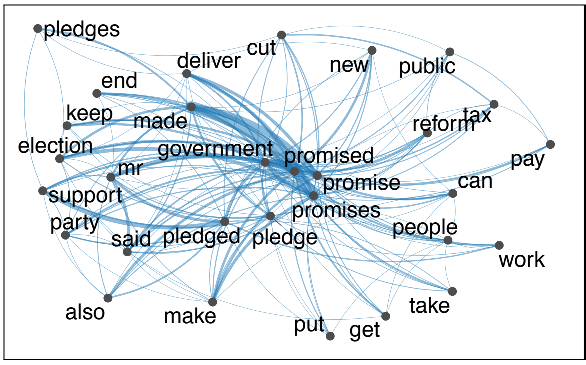
(a) Australia



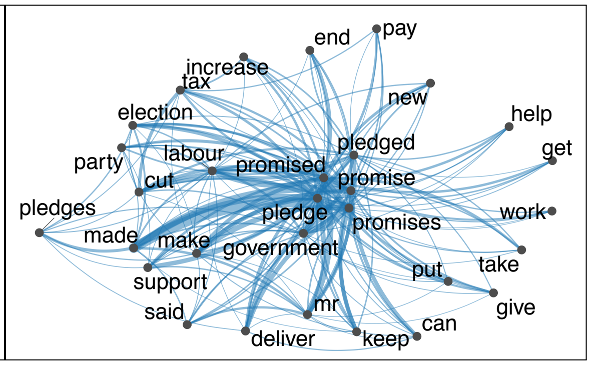
(b) Canada



(c) Ireland



(d) United Kingdom



C.3 Selection of Newspaper Articles

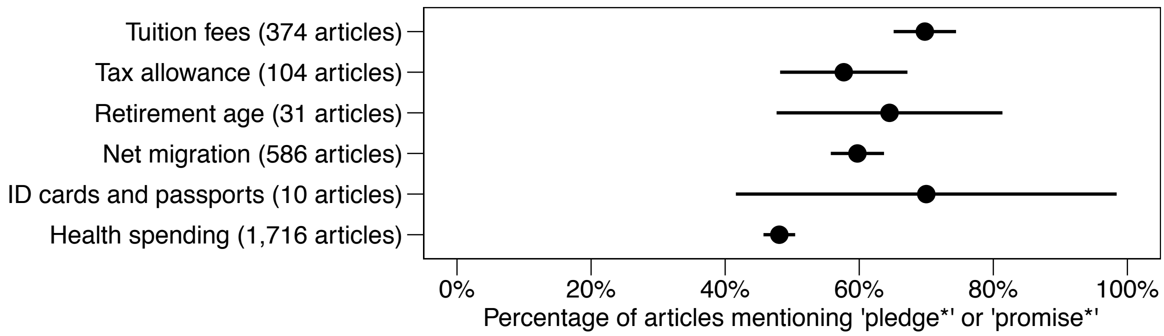
For this paper, I downloaded articles based on pledge-related keywords instead of actual pledges from manifestos. The question remains as to whether this search query systematically excludes articles that mention the progress or fulfillment of a campaign promise *without* using terms like `pledge*` or `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 & Brandenburg 2019). 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 Star, the Financial Times, the Independent, the Mail on Sunday, The Guardian, The Sun, and The Times. Table A4 lists the keywords used for each pledge. In a second step, I checked whether the article 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 me 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.

Table A4: Search query for retrieving articles about pledges mentioned in Thomson & Brandenburg (2019)

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

The corpus consists of 2,821 newspaper articles that mention one of the keywords from Table A4 and the name of a political party. Figure A8 plots the proportion of articles that contain pledge-related terms for each of the policies. Between 45% and 65% of the news articles include at least one word broadly indicating a promise (`promise*`, `pledge*`, `guarantee*`, `assure*`, or `ensure*`). This robustness check suggests that the text corpus used in this paper does not seem to exclude relevant reports about specific promises systematically. Therefore, it is reasonable to retrieve articles based on pledge-related keywords instead of retrieving newspaper articles based on terms relating to concrete policies.

Figure A8: Articles about policies used in Thomson & Brandenburg (2019) that include pledge-related terms



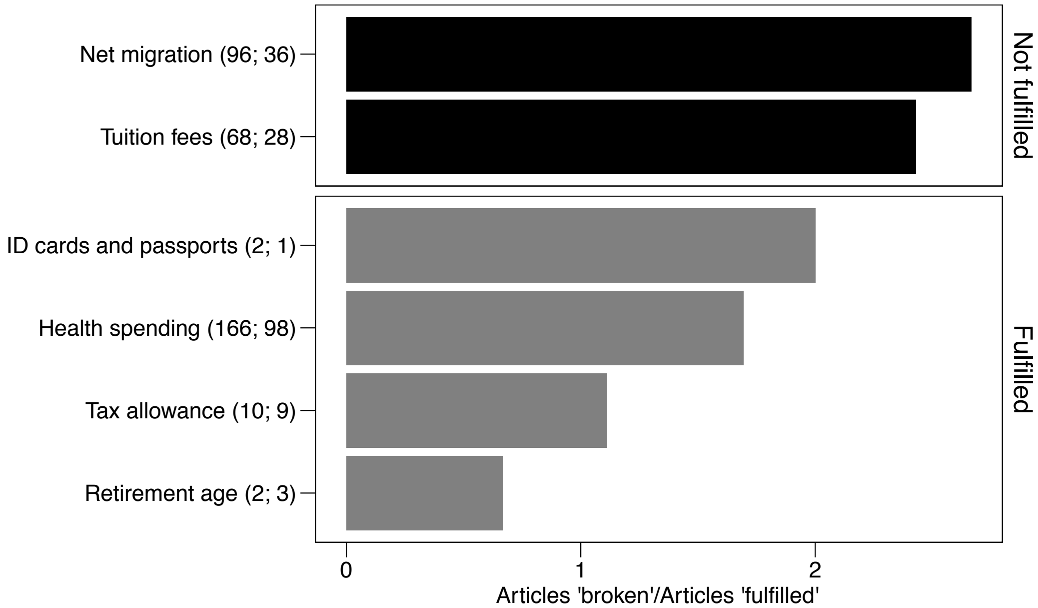
Note: The plot shows the proportion of articles from three UK newspapers published in the three months prior to the 2015 General Election, along with 95% confidence intervals.

Figure A9 plots the ratio of news articles focusing predominantly on the breaking of a promise to the articles focusing on fulfillment of the same promise. First, we observe that the two broken promises received more media attention than the fulfilled promises (except for the 'health' pledge). Secondly, the plot seems to suggest that even pledges which have been fulfilled in reality are not exclusively reported as being fulfilled in British newspapers. Yet, an extensive manual classification by human coders must be conducted to make inferences about the question as to whether the media fail to report on the fulfillment or breaking of pledges accurately. Instead, this validity check mainly concerns the question of whether the keyword-based analysis of promise-related content picks up the relevant news articles. This appears to be the case.

C.4 Assessing the Content of the Class 'Ongoing Promises'

Recall that sentences are classified as 'ongoing promises' if (1) the news article contains the name of a political party or refers to a manifesto, and if (2) the sentence contains a word indicating a promise as well as (3) one of the 75 most frequent verbs co-occurring with terms indicating a promise in party manifesto.

Figure A9: The ratio of articles classified as mentioning the breaking and fulfillment of a promise



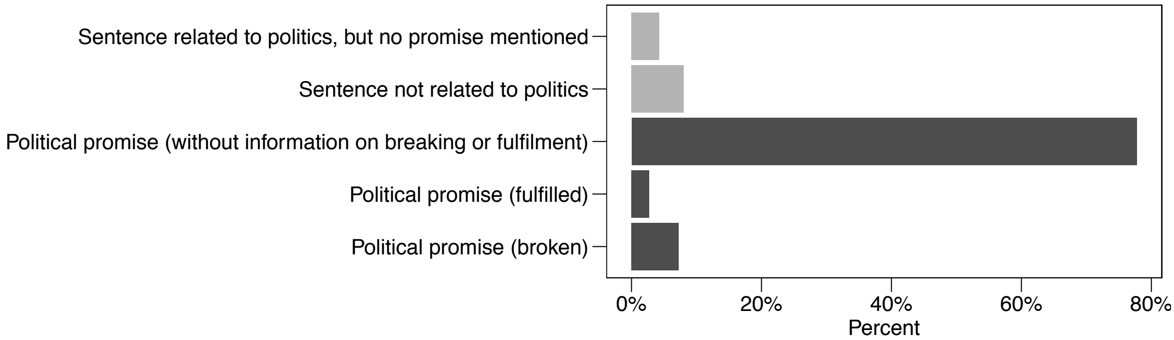
Note: The first number in brackets shows the absolute number of articles classified as 'broken', the second number indicates the number of articles classified as 'fulfilled'.

Since 'broken' and 'fulfilled' promises are measured using a rather small, but very concrete set of keywords (as suggested by Muddiman et al. 2019), the class 'ongoing' promises might contain statements on the breaking or fulfillment of promises which are not captured through the keywords used for the classification of broken and fulfilled promises. To test for this possibility, I randomly selected 400 sentences from the 'ongoing promises' class (100 sentences from each country) and hand-coded each sentence in the following way:

1. Does the sentence relate to politics?
2. Does the sentence contain information on a political promise?
3. Does the sentence mention the fulfillment or breaking of a promise?

Across the four countries, almost 80% of the sentences from the sample relate to political promises, but do not include any detailed information on the breaking or fulfillment of a promise (Figure A10). 8% of the sentences do not relate to politics; 4% of the sentences relate to politics but do not mention a promise. This result, along with the other validation tests reported in this paper, shows that the class 'ongoing promises' is meaningful and that the vast majority of sentences relate to political pledges. 7% of the sample contains information on the breaking of a promise; 3% of the statements include information on fulfilled promises. Yet, in almost all of these cases, the details on the fulfillment or breaking are very vague. This finding mirrors the crowd-sourced coding task (SI Section C.1): the dictionary approach performs well in separating promise-related content into the three categories of interest.

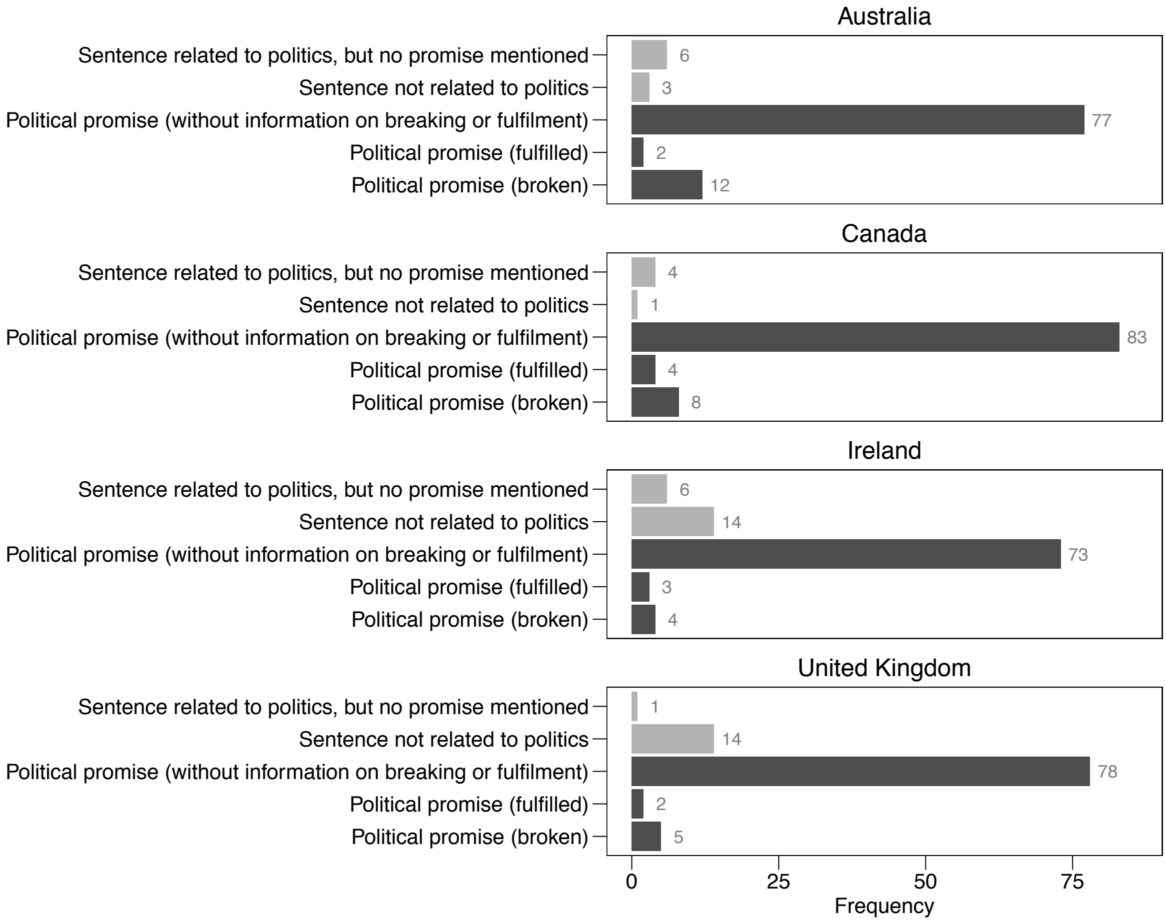
Figure A10: A manual inspection of the content of 'ongoing promises'



Note: Results are based on a hand-coded sample of 400 sentences (100 sentences from each country) that were classified as 'ongoing promises'.

Even though broken promises and fulfilled promises are rarely mentioned, the number of statements on broken promises (29) is almost three times higher than the number of statements on fulfilled promises (11). Figure A11 reports the results separately for each country. Overall, this validation exercise points towards the negativity bias in pledge reporting (see also SI Section E.2) and underscores that 'ongoing promises' is a meaningful and important category.

Figure A11: A manual inspection of the content of 'ongoing promises' (separately for each country)



Note: Results are based on a hand-coded sample of 400 sentences (100 sentences from each country) that were classified as 'ongoing promises'.

D Additional Plots and Tables

Tables A5–A7 show the absolute number and relative frequency of sentences classified as 'broken', 'fulfilled', and 'ongoing'. The distribution of sentences in each category for the primary classification is shown in the paper (Table 1). Sentences classified as 'other' (see SI Section B) are excluded from the analysis. Results shows that aggregated proportions are very similar for the two dictionaries and the two ways of measuring political promises (i.e., either taking all sentences in articles that contain the name of a political party, or only focusing on sentences and their immediate context of ±1 sentence that contain a name of an MP or a party). In all four cases, the proportion of 'broken' promises exceeds the proportion of 'fulfilled' promises by a considerable margin.

Table A5: Number of relevant sentences for each category and country (inductive, sample of statements directly referring to politician or party)

Country	Ongoing	Broken	Fulfilled	Total
Australia	45,268 (92.0%)	2,921 (5.9%)	1,006 (2.0%)	49,195
Canada	32,851 (93.5%)	1,494 (4.3%)	779 (2.2%)	35,124
Ireland	19,496 (93.1%)	895 (4.3%)	540 (2.6%)	20,931
United Kingdom	217,540 (95.0%)	7,710 (3.4%)	3,798 (1.7%)	229,048

Table A6: Number of relevant sentences for each category and country (deductive, full sample)

Country	Ongoing	Broken	Fulfilled	Total
Australia	53,114 (92.5%)	3,299 (5.7%)	977 (1.7%)	57,390
Canada	51,902 (92.9%)	2,218 (4.0%)	1,737 (3.1%)	55,857
Ireland	32,129 (93.6%)	1,304 (3.8%)	899 (2.6%)	34,332
United Kingdom	274,589 (94.9%)	8,274 (2.9%)	6,541 (2.3%)	289,404

Table A7: Number of relevant sentences for each category and country (deductive, sample of statements directly referring to politician or party)

Country	Ongoing	Broken	Fulfilled	Total
Australia	45,849 (93.2%)	2,661 (5.4%)	709 (1.4%)	49,219
Canada	32,965 (93.9%)	1,271 (3.6%)	873 (2.5%)	35,109
Ireland	19,745 (94.3%)	783 (3.7%)	404 (1.9%)	20,932
United Kingdom	218,701 (95.5%)	6,065 (2.6%)	4,247 (1.9%)	229,013

Table A8 contains the output of negative binomial regressions used to estimate the predicted counts of newspaper coverage in Figure 3. Figure A12 shows generalized additive models for pledge coverage for each class under single-party cabinets and multiparty cabinets. Figure A13 reproduces this plot but recodes all coalitions between the Liberal Party of Australia and the National Party of Australia as single-party cabinets. The shapes of the curves remain very similar. The shapes are similar to the predicted counts in Figure 3, offering support for the accuracy of the modeling choice and the conclusion that coverage of fulfilled promises does not have a first peak in the first half of an electoral cycle.

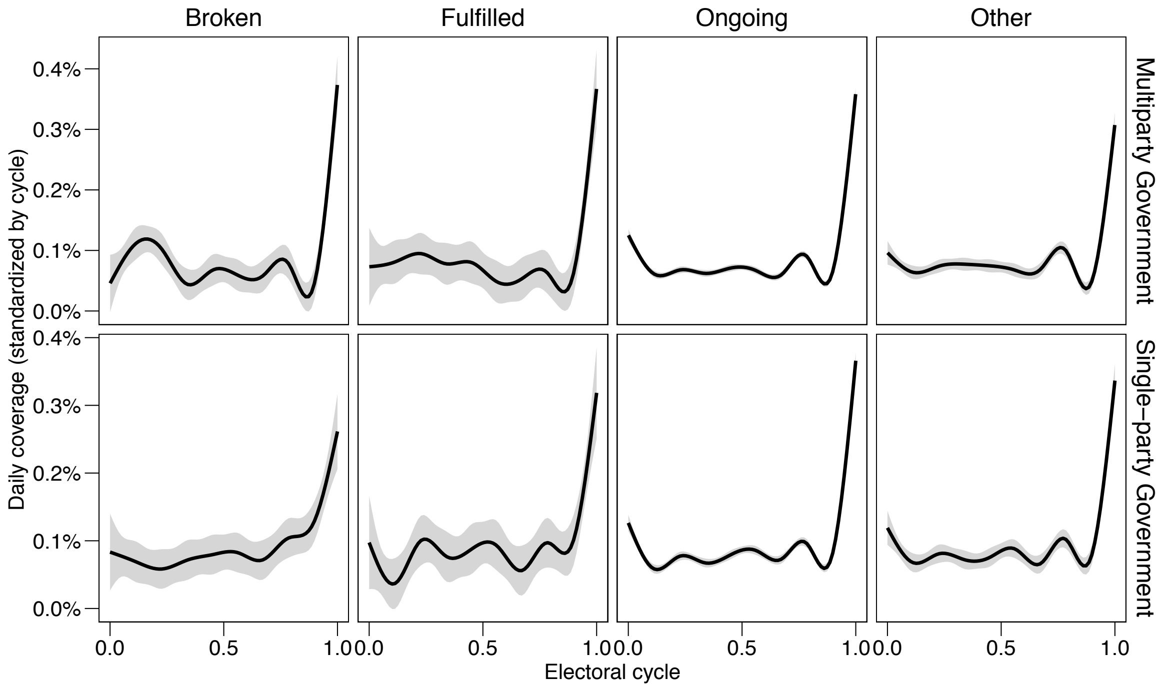
Table A8: Predicting the number of articles on promises published throughout the electoral cycle

	M1: Ongoing (single)	M2: Broken (single)	M3: Fulfilled (single)	M4: Ongoing (multi)	M5: Broken (multi)	M6: Fulfilled (multi)
(Intercept)	0.01	-3.81^{***}	-4.97^{***}	1.27^{***}	-2.02^{***}	-3.21^{***}
	(0.05)	(0.22)	(0.39)	(0.04)	(0.12)	(0.21)
Number of articles per class (lag)	0.02^{***}	0.26^{***}	0.40^{***}	0.02^{***}	0.31^{***}	0.26^{**}
	(0.00)	(0.02)	(0.04)	(0.00)	(0.02)	(0.09)
Electoral cycle	21.27^{***}	39.03^{***}	20.90^{***}	15.63^{***}	5.12	15.66^{*}
	(1.26)	(4.39)	(6.21)	(1.34)	(4.81)	(7.52)
Electoral cycle^2	20.94^{***}	19.30^{***}	24.60^{***}	19.87^{***}	26.53^{***}	30.49^{***}
	(1.16)	(3.96)	(5.46)	(1.18)	(4.12)	(6.41)
Electoral cycle^3	12.11^{***}	22.12^{***}	25.32^{***}	8.37^{***}	29.24^{***}	28.29^{***}
	(1.09)	(3.77)	(5.15)	(1.14)	(4.01)	(6.11)
Lagged GDP change (log)	-0.01	-0.07	-0.17^{*}	-0.04^{*}	0.16^{*}	-0.04
	(0.01)	(0.06)	(0.08)	(0.02)	(0.07)	(0.11)
Poll change to previous election	-0.01^{***}	-0.02^{**}	-0.02^{**}	0.00	0.01	0.03
	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)	(0.02)
AIC	152287.14	21974.82	11446.54	79719.46	10653.84	4809.22
BIC	152525.76	22213.44	11685.17	79857.28	10791.65	4947.03
Log likelihood	-76114.57	-10958.41	-5694.27	-39841.73	-5308.92	-2386.61
Deviance	32264.33	8257.00	4928.86	17983.52	4178.00	2094.91
N	27678	27678	27678	15620	15620	15620

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

Note: Models 1–3 limit the sample to single-party cabinets, Model 4–6 focus only on multiparty cabinets. All models include dummy variables for each country-cycle observation which are omitted from the table. Standard errors in parentheses.

Figure A12: The daily proportion of pledge-related statements throughout the electoral cycle under single-party governments and multiparty cabinets

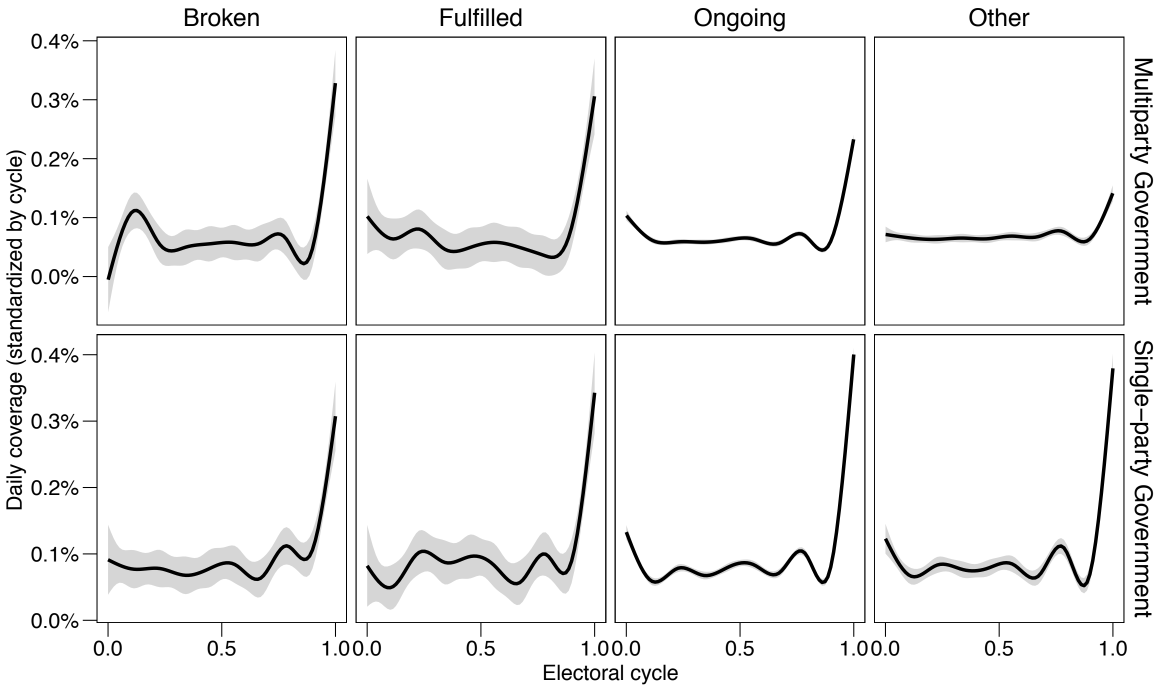


Note: The lines are based on generalized additive model (GAM) smoothers. The dates are standardized from 0 to 1 for each cycle, where 0 marks the date of *Election t*, and the date of *Election t+1* is set to 1. This plot codes the Liberal Party of Australia and the National Party of Australia as two distinct parties.

Figure A14 shows the predicted values of the focus on broken relative to fulfilled promises for broadsheets and tabloids, mirroring the findings from Figure 5. The regression suggests an increase in the focus of broken promises relative to fulfilled promises over time. I also use the raw data and plot loess regression lines for the relationship between year and focus on broken promises (Figure A15). The plot suggests that the conclusions from the regression analysis do not depend on the modeling choice (treating the year as a continuous variable). Despite some fluctuations, we observe moderate to strong increases of coverage of broken relative to fulfilled promises in the four countries.

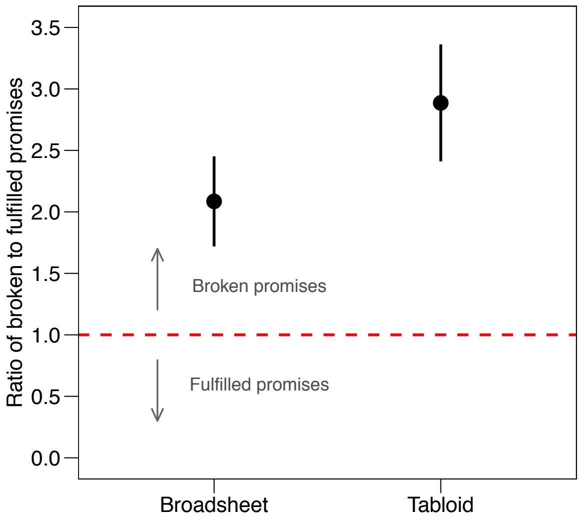
Figure A17 plots the yearly proportions of statements published in broadsheets. Statements from broadsheets exceed statements from tabloid newspapers in all four countries. In addition, the share of statements from broadsheets has increased over time, suggesting that the higher focus on broken promises is not related to an increasing number of statements on promises in usually more negative tabloid papers.

Figure A13: The daily proportion of pledge-related statements throughout the electoral cycle under single-party governments and multiparty cabinets (with a different measurement of cabinets in Australia)



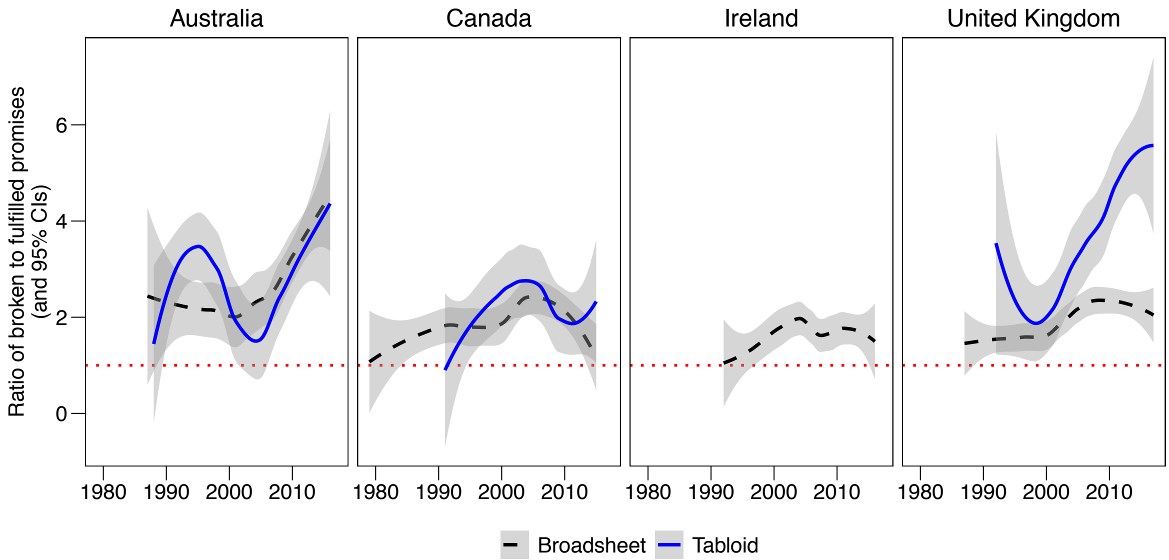
Note: The lines are based on generalized additive model (GAM) smoothers. The dates are standardized from 0 to 1 for each cycle, where 0 marks the date of *Election t*, and the date of *Election t+1* is set to 1. This plot codes the Liberal Party of Australia and the National Party of Australia as the same party. As a result, all electoral cycles in Australia are coded as single-party governments.

Figure A14: Predicting the ratio of reports on broken to fulfilled promises per quarter and newspaper



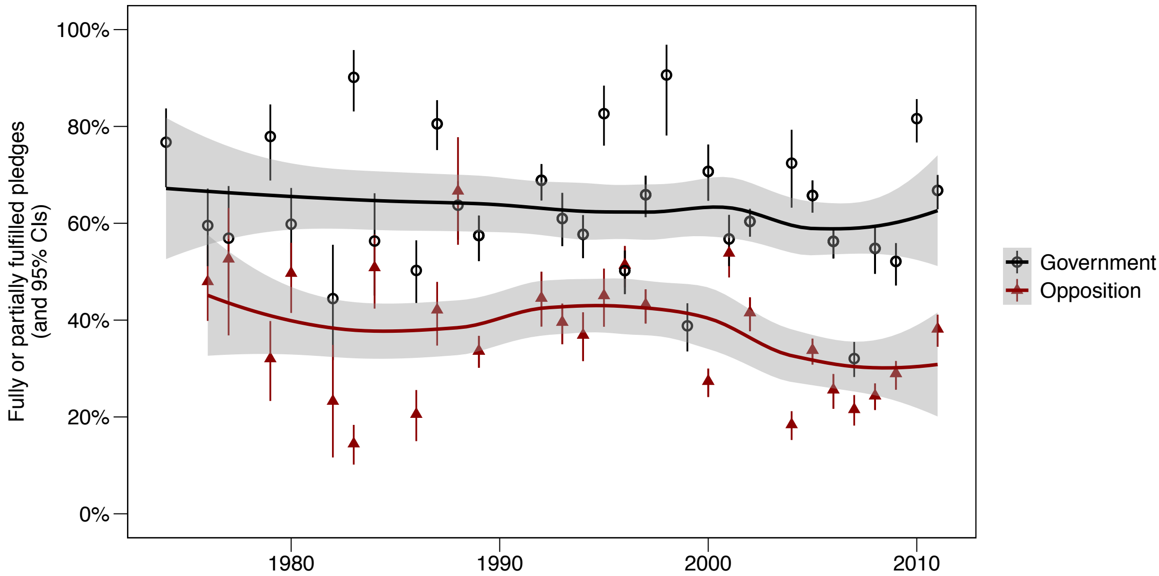
Note: The estimates are based on Model 1 of Table 2. A ratio of 1 implies that the number of statements on broken promises equals the number of statements on fulfilled promises. Vertical bars and shaded areas show 95% confidence intervals.

Figure A15: The quarterly focus on broken promises relative to fulfilled promises over time



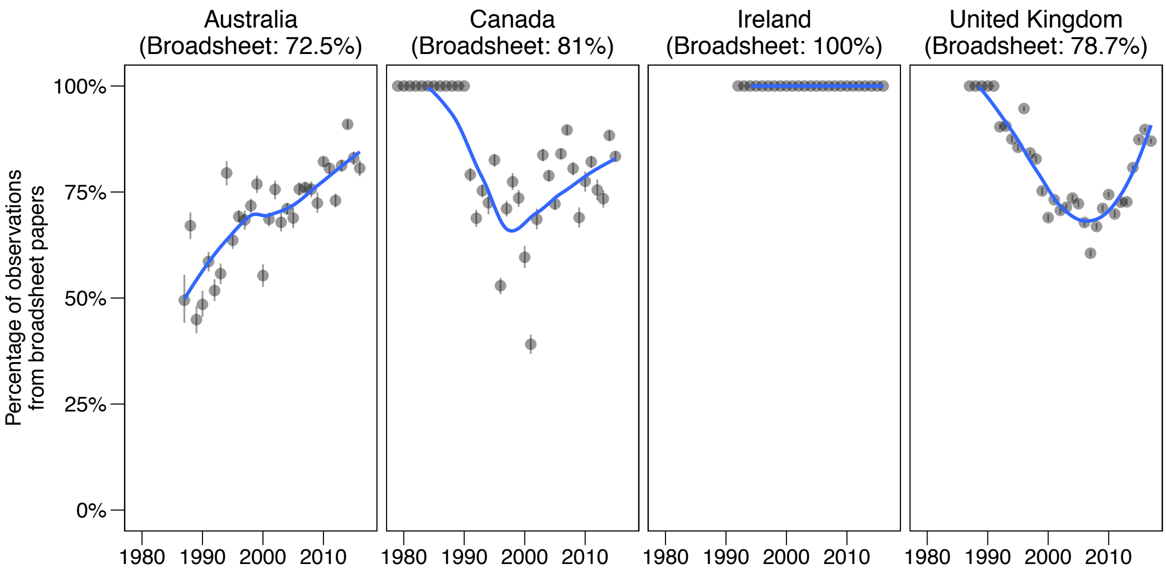
Note: Loess regression lines for broadsheet and tabloid newspapers in each country. Grey areas indicate 95% confidence intervals.

Figure A16: The proportion of fully or partially fulfilled pledges by government and opposition parties



Note: A reanalysis of the data provided by Thomson et al. (2017). Points and errorbars show the average pledge fulfillment per year along with 95% bootstrapped confidence intervals. The analysis includes 12 countries.

Figure A17: The percentage of statements from broadsheet papers relative to tabloids



Note: The vertical bars indicated 95% bootstrapped confidence intervals.

E Robustness Tests

E.1 Different Dictionary and Alternative Aggregation

This section shows the regression results and plots with predicted values using an alternative dictionary, which was developed deductively from the comparison to human coding. Importantly, the results do not change.

Table A9: Predicting the ratio of broken to fulfilled promises by quarter and newspaper based on the deductively developed dictionary

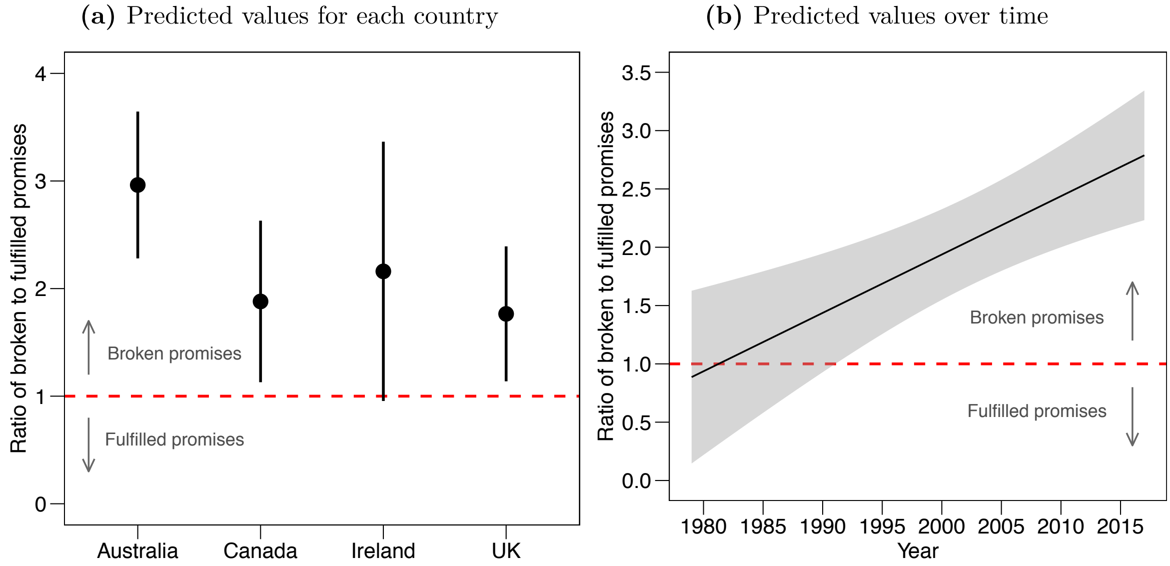
	M1: Full sample	M2: Logged DV
Newspaper: Tabloid	0.79^{*}	0.27^{**}
	(0.33)	(0.09)
Canada	-1.08^{*}	-0.39^{*}
	(0.53)	(0.17)
Ireland	-0.80	-0.19
	(0.69)	(0.22)
United Kingdom	-1.20^{*}	-0.49^{**}
	(0.48)	(0.16)
Lagged GDP change (log)	0.11	0.04
	(0.14)	(0.04)
Poll change to previous election	-0.02	-0.01^{*}
	(0.01)	(0.00)
Year	0.05^{***}	0.03^{***}
	(0.01)	(0.01)
Gov. type: Single-party government	0.44	0.16
	(0.37)	(0.14)
AIC	6720.21	3238.11
BIC	6784.05	3301.95
Log likelihood	-3348.10	-1607.05
N	1510	1510
N (Cycles)	32	32
N (Newspapers)	22	22

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.
Note: Model 1 uses the full sample of quarters with at least one sentence on broken and one sentence on fulfilled promises. Model 2 uses a logged dependent variable of the ratio. Models include random intercepts for each newspaper and cycle. Intercepts omitted from table. Standard errors in parentheses.

Figure A18 shows the predicted values reported in the main paper, but after calculating the ratio of broken to fulfilled promises using the deductive dictionary. The results remain unchanged.

Figure A19 plots the correlation between the aggregated logged focus on broken relative to fulfilled promises for the inductive and deductive dictionary. Each point displays the

Figure A18: Predicting the ratio of reports on broken to fulfilled promises per quarter and newspaper using a deductively developed dictionary

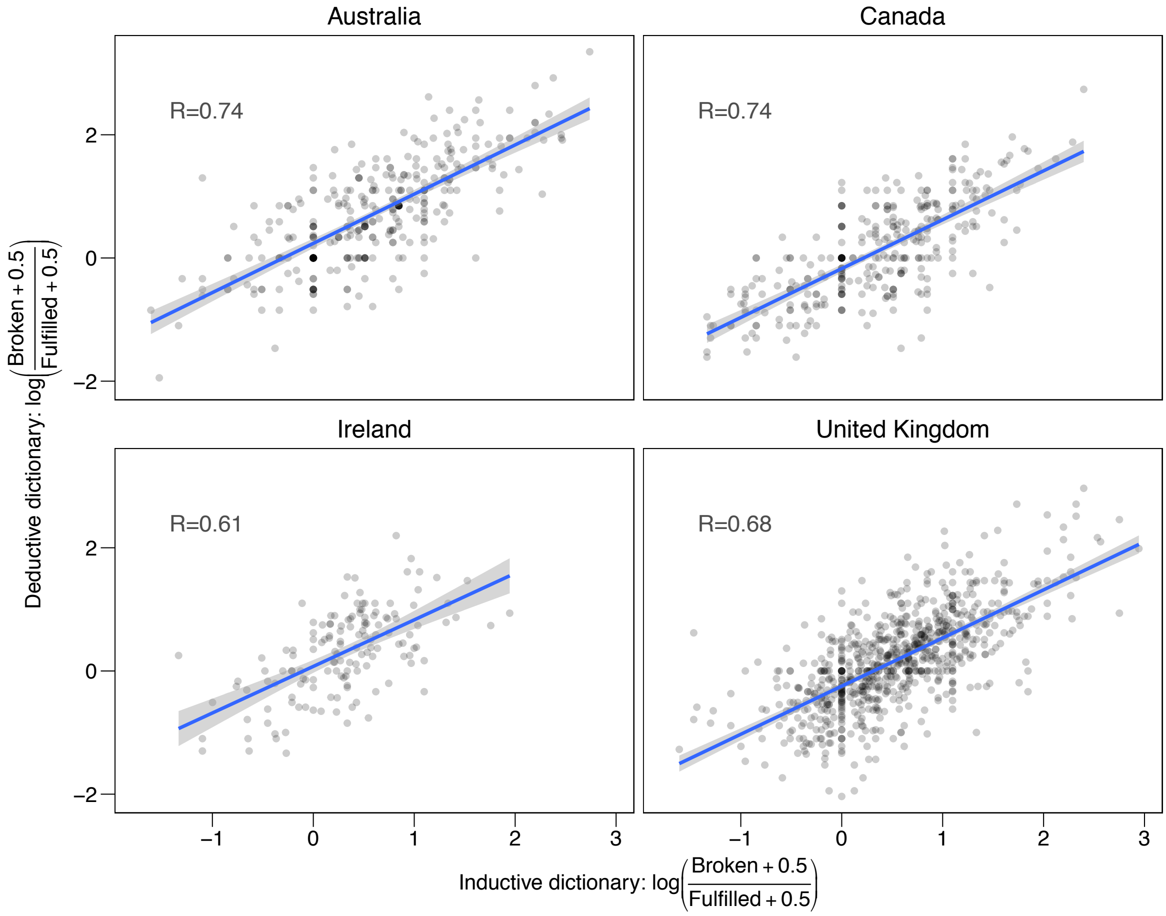


Note: The estimates are based on Model 1 of Table A9. A ratio of 1 implies that the number of statements on broken promises equals the number of statements on fulfilled promises. Vertical bars and shaded areas show 95% confidence intervals.

proportion of all articles published by one newspaper in a given quarter. Correlations between both dictionaries range from 0.61 to 0.74.

Finally, I estimate predicted values for the logged ratio of broken to fulfilled promises to account for outliers and avoid skewness of the dependent variable (Lowe et al. 2011). Now, a value above 0 implies a higher focus on broken relative to fulfilled promises. In the four countries, we observe a consistently higher coverage of broken promises relative to fulfilled promises (Figure A20).

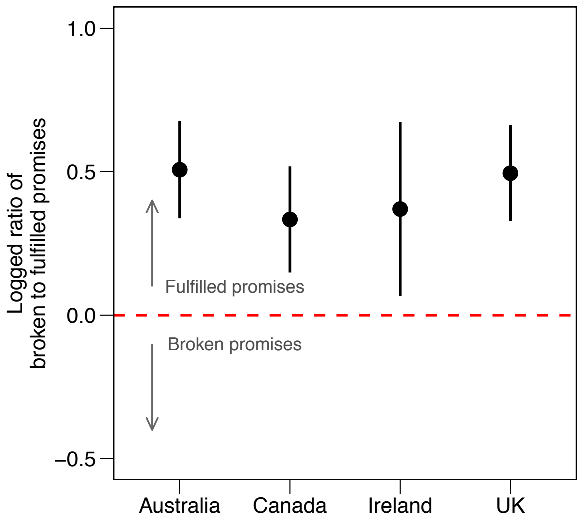
Figure A19: Comparing the focus on broken and fulfilled promises, using two dictionaries



Note: The values on the x-axis show the proportion of broken and fulfilled promises using a dictionary that only relies on verbs in the past tense; the values on the y-axis show the proportion using the keywords that maximize F1 scores when comparing sentence-level classifications to human coders

.

Figure A20: Predicting the ratio of reports on broken and fulfilled promises per quarter and newspaper using a logged measure of the ratio of broken to fulfilled promises



Note: The estimates are based on Model 3 of Table 2. A value of 0 implies that the number of statements on broken promises equals the number of statements on fulfilled promises. Vertical bars and shaded areas show 95% confidence intervals.

E.2 Sentiment Analysis

The proportions of statements on broken and fulfilled promises could rely on the selection of keywords used for the classification (Muddiman et al. 2019). Therefore, I also conduct a sentiment analysis to test whether news outlets exhibit a negativity bias when reporting on promises. The process works as follows. First, I remove all terms that are part of the pledge dictionary (to avoid that some terms of the dictionary contribute to the sentiment). Afterwards, I apply the Lexicoder Sentiment Dictionary, a collection of positive and negative terms (and their respective negations) which has been validated carefully for political text and economic news (Young & Soroka 2012; Proksch et al. 2019). Third, after having applied the dictionary to the tokenized text, I aggregate sentiment for each sentence using the formula recommended by Proksch et al. (2019):

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

Next, I estimate the quarterly sentiment for all sentences classified either as broken or fulfilled, and compare the values with the 'baseline' of sentences classified as ongoing promises in the same newspaper. Lower values imply more negative tone. If pledge coverage was balanced, we would expect sentiment of sentences on broken and fulfilled promises to be at least similar to the baseline of 'ongoing' promises. Figure 6 in the main paper reveals that sentences classified *either* as broken or fulfilled are substantively more negative than the baseline category.

To extend the analysis to the level of sentences and include control variables, I run multilevel linear regressions with random effects for cycles and newspapers (Table A10). The main independent variable is the *Class*. If sentiment is negative for the classes of broken or fulfilled promises, media outlets tend to exert a negativity bias when mentioning promises. I also control for the type of newspaper (tabloid vs. broadsheet), the year, as well as changes in the economic situation and public approval of the government party/parties. Again, sentiment is substantively less positive in statements on broken or fulfilled promises compared to the baseline sentiment, both for the inductively and deductively derived dictionary classification. The coefficient of 0.62 corresponds to 60% of the standard deviation of the dependent variable.

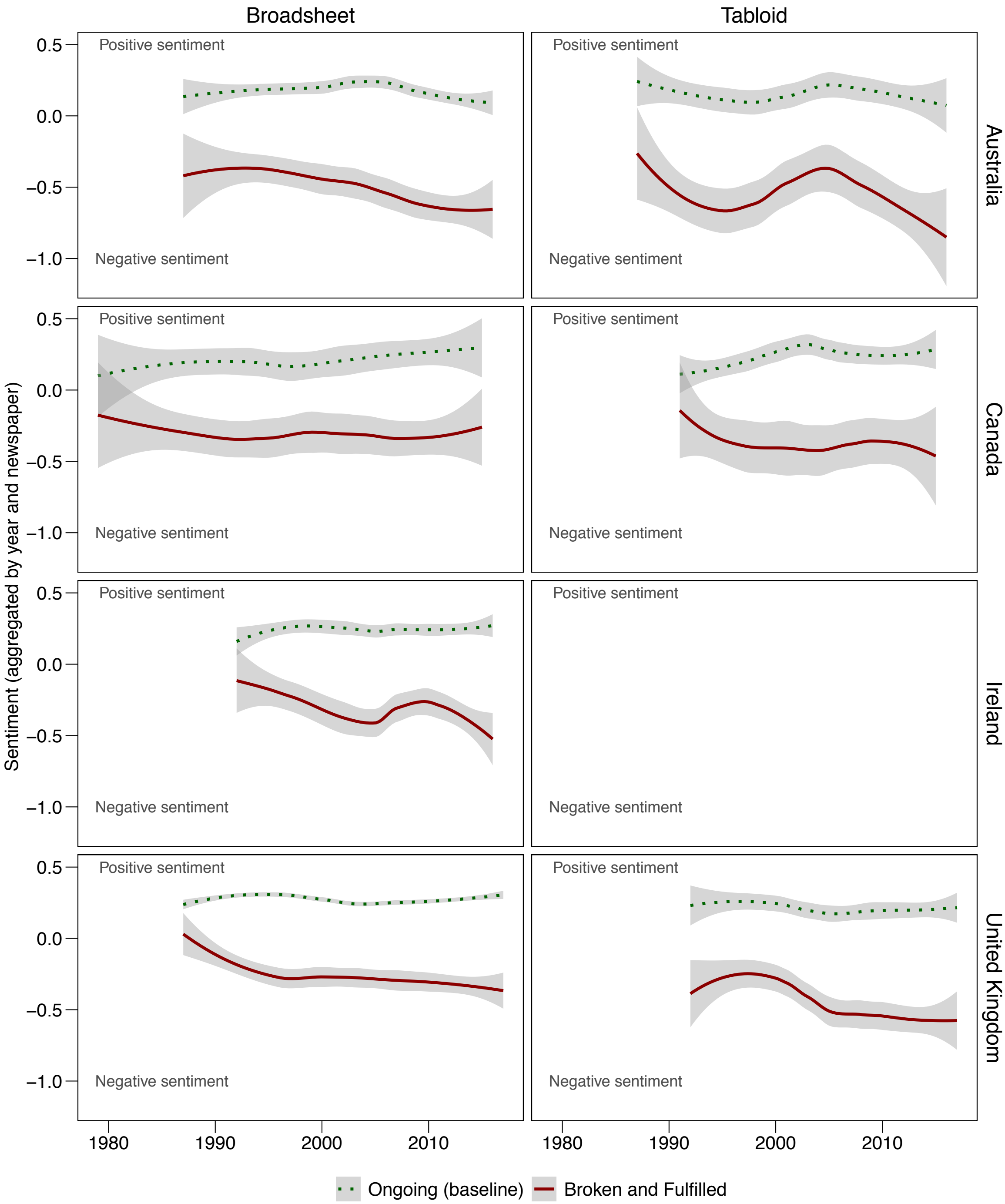
Table A10: Predicting sentiment in sentences about political promises

	M1: Inductive	M2: Deductive
(Intercept)	-25.36^{***}	-24.27^{***}
	(2.48)	(2.47)
Class: Broken or Fulfilled	-0.62^{***}	-0.56^{***}
	(0.01)	(0.01)
Newspaper: Tabloid	-0.04^{**}	-0.04^{**}
	(0.01)	(0.01)
Lagged GDP change (log)	-0.01^{*}	-0.01^{**}
	(0.00)	(0.00)
Poll change to previous election	0.00^{***}	0.00^{***}
	(0.00)	(0.00)
Year	0.01^{***}	0.01^{***}
	(0.00)	(0.00)
Gov. type: Single-party government	0.04	0.04
	(0.05)	(0.05)
AIC	1173226.45	1174174.46
BIC	1173335.63	1174283.63
Log likelihood	-586603.23	-587077.23
N	407696	407123
N (Cycles)	32	32
N (Newspapers)	22	22

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.
Note: Model 1 uses the inductive classification of fulfilled, broken, and ongoing promises, Model 2 the deductive classification. Both models include random intercepts for each cycle and newspaper. Standard errors in parentheses.

Finally, I plot loess regression lines that estimate the development of sentiment per outlet and quarter over time (Figure A21). I distinguish between broadsheets and tabloids. While the baseline sentiment (green dashed line) remains relatively stable, in tabloids and broadsheets (with the exception of broadsheets in Canada) we observe a slight increase in negativity over time. This visual evidence provides further support for the hypothesis that pledge coverage has become more negative over time.

Figure A21: The development of sentiment in statements on ongoing promises and on broken and fulfilled promises over time



Note: Lines show smoothed loess regression lines of sentiment (and 95% confidence intervals), aggregated for each newspaper and year. The sample includes only Irish broadsheet newspapers.

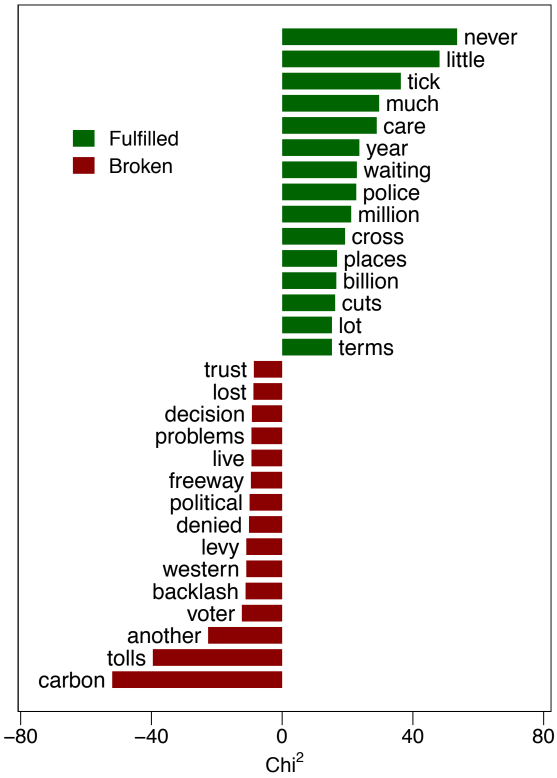
E.3 Identifying Important and Distinct Terms

I identify important and more 'unique' terms in the classes 'broken' and 'fulfilled' for newspapers from each country using keyness statistics. Keyness is a signed two-by-two association which identifies frequent words in documents in a target and a reference group (Benoit et al. 2018). The target group is 'fulfilled', the reference group are sentences classified as 'broken'. The chi-square value for a term is positive (negative) when the observed value for a term is higher (lower) than the expected value. Before estimating keyness for each country, proper nouns, numbers, and terms with fewer than three characters are removed. The results of the keyness analysis in Figure A22 allow for two conclusions. First, the terms with very high and low values correspond to our intuition. In the United Kingdom, `tuition` and `fees` are very often used in sentences on broken promises, which is reasonable since the central promise by the Liberal Democrats in 2010 of not increasing tuition fees had been broken and received a lot of public attention during the legislative cycle. Other terms, such as `accused`, `anger`, `angry`, `betrayed`, `lost`, `scandal`, `taxes` and `tired` are terms we would expect to occur in statements on broken pledges.

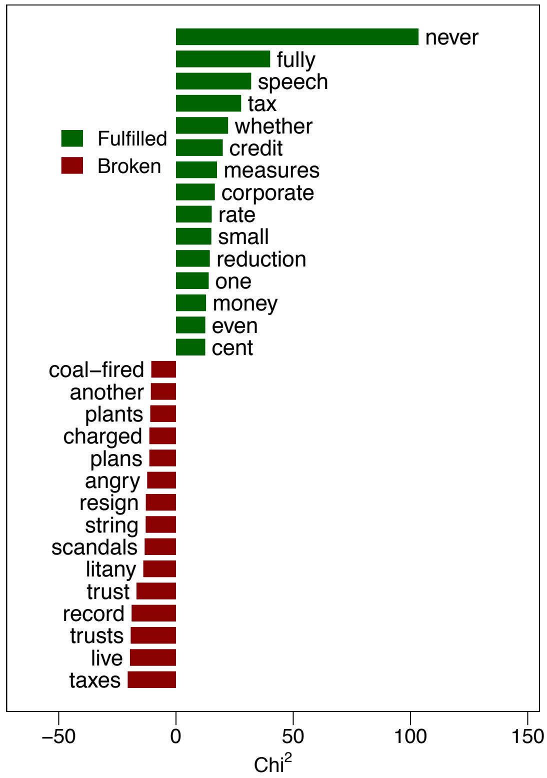
Second, some of the terms more unique to 'fulfilled' promises tend to exhibit a negativity bias. For example, the term `never` is most unique for 'fulfilled' promises in all four countries, suggesting that politicians or journalists express concerns that a promise has *not* been fulfilled. Terms like `little`, `small` also appear more frequently in the 'fulfilled' category in some of the countries. This qualitative assessment and the sentiment analysis (SI Section E.2) seem to suggest that the findings presented in the paper can be interpreted as conservative estimates of the negativity bias in media coverage of political promises.

Figure A22: Keyness plots with important terms for sentences classified as fulfilled and broken

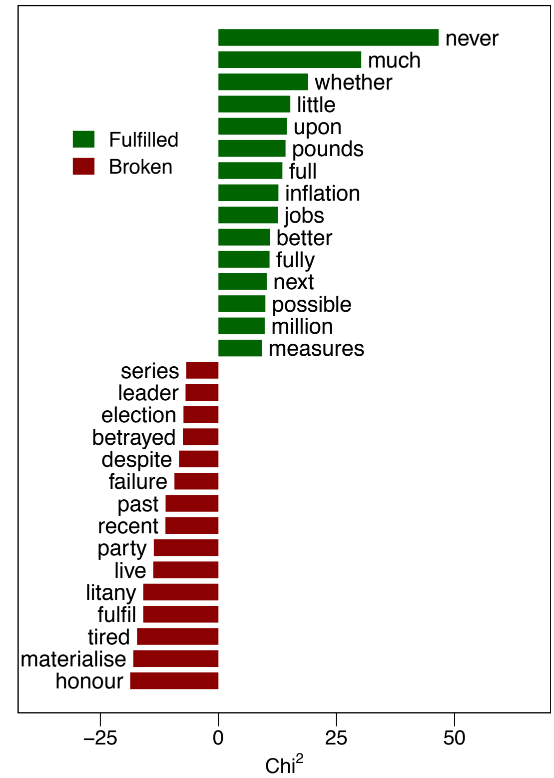
(a) Australia



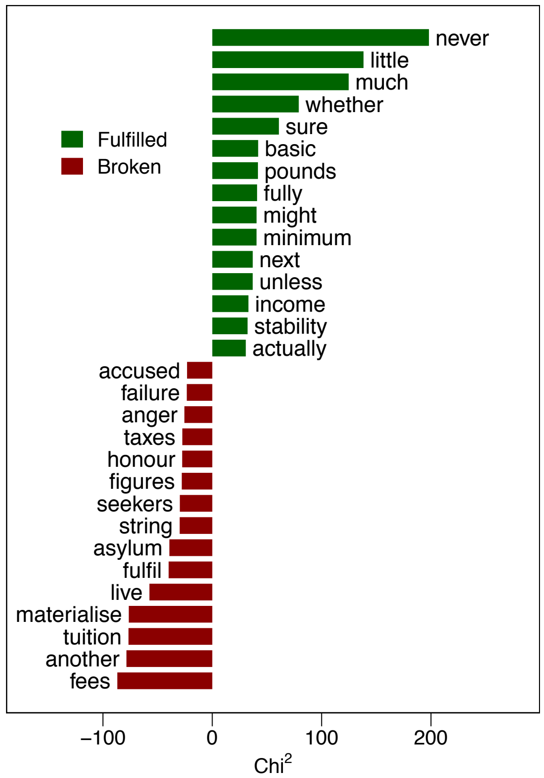
(b) Canada



(c) Ireland



(d) United Kingdom



F 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 ongoing 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.