Does Public Opinion Affect Elite Rhetoric?

July 3, 2019

Abstract

Does public opinion affect elite rhetoric? This central question of political science has received little empirical scrutiny. Of particular interest is whether public opinion affects i) what topics elites address and ii) what positions they endorse. We add to this debate by drawing on unique evidence from Germany. In 2015, a legal ruling forced the German government to declassify all its public opinion research. Our causal identification strategy exploits the demonstrably exogenous timing of the reports' dissemination to cabinet members within a window of a few days. We find that exposure to the public opinion reports leads elites to change their rhetoric markedly. Specifically, linguistic similarity between elite speech and public opinion increases significantly after reports are disseminated—a finding that points toward rhetorical agenda setting. By hand-coding a subset of 2,000 report-speech pairs, we also demonstrate that elites substantively adapt their rhetoric to majority opinion.

Keywords: Public opinion; Elite rhetoric; Text analysis; Responsiveness

Word count: 9,998

1 Introduction

Does public opinion affect elite rhetoric? This central question of political science has received little empirical scrutiny. The handful of published studies have yielded conflicting findings. One influential study by Jacobs and Shapiro (2000) finds no evidence that public opinion meaningfully affected the rhetoric of President Clinton. By contrast, Rottinghaus (2008) uses White House archival data and argues that public opinion places significant constraints on presidential framing. Evidence that reliably establishes a causal connection between public opinion and elite rhetoric, however, is not existent.

What explains the lack of evidence? First, elected officials typically mask their means of gauging public opinion and their exposure thereto. Second, the dissemination of public opinion to elites must be exogenous if one wants to establish a causal relationship. Third, when analyzing elite rhetoric researchers face nontrivial measurement challenges. Any analysis into the interplay of public opinion and elite rhetoric must differentiate whether it explores i) the rhetorical agenda elites choose, or ii) the substantive positions they endorse.

The present article circumvents these challenges by using an unusual source of information: classified governmental public opinion research. In 2013, a politician of the German Green party sued the German government in order to gain access to its classified public opinion research. The German Federal Press Office (*Bundespresseamt*, henceforth BPA) subsequently and reluctantly granted the politician access to all public opinion research conducted between 2009 and 2013. Overall, the politician, together with a team of journalists, hand copied more than 10,000 pages of classified public opinion research.

These research reports offer a unique lense into elite exposure to public opinion. They are addressed directly to Germany's chancellor Angela Merkel and are disseminated to all cabinet members. The purpose of the reports is to give political elites a succinct overview of German public opinion. The research is conducted by Germany's major public opinion research firms. Most reports include up to 60 pages of qualitative and quantitative insights.

It is here that cabinet members find a rich repertoire to attune, or not, their rhetoric to public opinion.

How does exposure to public opinion reports affect elite rhetoric? We hypothesize that exposure to public opinion can lead to two different rhetorical reactions. First, it may lead elites to change their rhetorical agenda (John and Jennings 2010; Mortensen et al. 2011). The logic, here, is that the reports make specific topics cognitively salient, and signal what topics are of relevance to the electorate. Second, exposure to public opinion may lead elites to adjust their substantive positions to the median voter (Eggers and Spirling 2014). After all, the research reports clearly communicate what the population thinks.

To explore these hypotheses, we take advantage of the fact that the reports indicate the date at which the BPA sent the research reports to the cabinet. And there are good reasons to believe that the precise timing is exogenous (within a window of a few days). The reports undergo a tedious tendering process and we show that the dissemination dates do not cluster around salient events such as elections or parliamentary sessions. The dissemination timing is also orthogonal to the media salience of a given reports' topic as well as to citizen satisfaction with the government.

Based on the plausibly exogenous timing of the reports' dissemination to cabinet members, our empirical strategy is to compare elite speech to the public opinion reports right before and right after a report was issued. We measure elite rhetoric using all published government speech documents from 2005 to 2016 ($>$ 20,000). To capture rhetorical agenda setting, we measure linguistic (cosine) similarity between the reports and elite speech. To explore substantive responsiveness, we measure agreement between the reports and elite speech by hand-coding a random subset of 2,000 speech-report pairs.

Using a regression discontinuity design, we yield two key pieces of evidence. First, we find that linguistic similarity between elite speech and the public opinion reports increases by 0.014 points (S.E.=0.007) on a zero-to-one scale after reports are passed on to the cabinet.

We interpret this finding to mean that exposure to public opinion affects elites' rhetorical agenda. Second, we find that substantive agreement increases by 0.2 points (S.E.=0.07) on a 7-point scale. The finding thus implies that exposure to public opinion also leads elites to become more responsive to the public's preferences.

Can our empirical strategy tell agenda setting and substantive responsiveness apart? To address the concern that substantive agreement overlaps with cosine similarity, we present two key pieces of evidence. First, we show that both measures are not correlated. Second, we demonstrate that effect heterogeneity by topic salience is different across the two measures. When elites address salient topics, we only see increases in cosine similarity, while substantive agreement does not change. By contrast, when elites address non-salient topics, we see increased substantive agreement, but no changes in cosine similarity. The finding helps clarify when elites use agenda setting and when they decide to reposition: When topics are salient, substantive realignment may be too risky, which makes agenda setting more attractive. The reverse holds for non-salient topics. Here, agenda setting carries little weight, but substantive repositioning may win over voters without risking elites' reputation.

Our results contribute to three related debates in political science. First, we add to a growing literature analyzing elite rhetoric (Binzer Hobolt and Klemmensen 2008). Scholars have long argued that elite rhetoric, particularly that of U.S. presidents, influences public opinion (Cohen 1999). We complement this literature by showcasing that causality also runs in the opposite direction. While others have explored this relationship (Rottinghaus 2006; Wood and Lee 2009), we are the first to provide causal estimates.

Second, our evidence adds to a debate whether executive leaders follow centrist or partisan considerations. By highlighting that German politicians adjust their rhetoric to public opinion, i.e., the median voter, we support a centrist reading of executive political behavior (e.g. Canes-Wrone 2006). By the same token, our evidence rejects a partisan reading, whereby executives cater to their own political clientele (Wood 2009).

Last, we contribute to a growing literature that applies text-analytical methods to political science (Grimmer and Stewart 2013; Lowe et al. 2011; Lucas et al. 2015; Roberts et al. 2014; Slapin and Proksch 2008). Recent empirical studies have assessed, i.a., bureaucratic transcripts (Egesdal, Gill and Rotemberg N.D.), and diplomatic cables (Gill and Spirling 2015). Analyses regarding internal government reports, however, are few and descriptive in nature. We bring large-N causal evidence to this literature, showing that public opinion affects elites' rhetorical agenda and substantive positions.

2 Theoretical Background

Against the backdrop of a large literature underlining the rhetorical craftsmanship of elected officials (see, Chong and Druckman 2007), little is known about the ways in which public opinion affects elite speech. Vaughn and Villalobos, for instance, lament that social scientists have "ignored the determinants of what the president actually says" (2006, 681). Shapiro and Jacobs second that "[w]e are only now beginning to learn about the relationship between presidents and the polling and public opinion analysis that has gone on in presidential administrations since the 1960s" (2001, 151). What is more, the few empirical accounts that assess whether elected officials adjust their rhetoric to public opinion are inconclusive.

On the one hand, scholars have presented evidence that questions whether public opinion has any influence on elite rhetoric. In one historical study, Hall (2002) uses interviews and archival data to assess whether President George W. Bush used public opinion research to forge his rhetoric. The author finds that Bush had a distinct distaste for polls, labeling them "phony and artificial" (Hall 2002, 531). Similarly, Jacobs and Shapiro (2000) analyze President Clinton's failed health care reform campaign and Newt Gingrich's "Contract with America" and find no evidence for rhetorical pandering. One rare large-N study is provided by Wood and Lee (2009) who develop a measure of presidential liberalism from 1945 to 2005 and find presidents rather unresponsive to mass political preferences.

On the other hand, a different set of scholars has presented evidence demonstrating that public opinion does affect elite rhetoric. They point out the skillful employment of public opinion research in several U.S. administrations. Rottinghaus (2008), for instance, uses White House archival data, including internal polling reports, and finds that public opinion placed serious constraints on presidential framing of foreign policy. Relatedly, Geer (1996) finds that the professionalization of the White House public opinion polling led presidents to adjust their rhetoric to public preferences. In one systematic empirical study, Rottinghaus (2006) matches presidential statements spanning nine administrations to public opinion polling and demonstrates robust congruence.

A key limitation of the existing empirical studies, however, is that they merely correlate elite rhetoric with public opinion. As such, they fall short of providing causal estimates and simply reiterate the well-established finding that public opinion correlates with elite speech. The prime reason for this shortcoming is that it is difficult to make a convincing case that elite exposure to public opinion is exogenous. In addition, we lack empirical evidence from outside the U.S. or even the oval office.

At a more fundamental level, it is also unclear *how* public opinion may affect elite rhetoric. Providing elected officials with new information about public opinion may lead to two distinct rhetorical reactions. First, obtaining up-to-date public opinion may inspire elites to put a given issue on their rhetorical agenda. Second, information on public opinion may lead elites to substantively align their rhetoric with the median voter.

2.1 Agenda setting

At a basic level, exposure to public opinion may lead elected officials to change their rhetorical agenda. We define such agenda setting as "the priority given to an issue" (Chong and Druckman 2007, 112). Put more simply, agenda setting captures whether an elected official talks about a specific issue or not. Why may exposure to public opinion affect elected officials'

rhetorical agenda? Two reasons are noteworthy.

First, the provision of public opinion may signal to elites that a given topic is of relevance to citizens. Public opinion research—including the reports commissioned by the BPA—tends to focus on salient issues. Public opinion is also known to be less volatile for salient issues (Weaver 1991). Thus, if a cabinet member is exposed to public opinion, she may interpret this as evidence that voters care about the topic, leading her to put the topic on her agenda.

Second, the provision of public opinion on a given topic makes the topic cognitively salient to the politician. And, like ordinary citizens, elites try to "minimize the cognitive burdens for forming judgments by drawing on those considerations that are most accessible" (Koch 1998, 211). Thus, if a cabinet member reads a report on tax reform, she may draw on this very topic in her next speech or newspaper interview because the report made the topic cognitively salient.

- **Hypothesis 1 (*Agenda setting*)**: Exposure to public opinion on a given topic makes elites more likely to put the topic on their rhetorical agenda

2.2 Responsiveness

Besides affecting elites' rhetorical agenda, exposure to public opinion may also lead politicians to change how they substantively address a given topic. More specifically, exposure to public opinion may spark (rhetorical) responsiveness—a core tenet of representative democracy. Here, we define responsiveness as "politicians follow[ing] preferences as they change" (Wlezien 2004, 2). Why may exposure to public opinion lead elected officials to endorse substantive positions in line with the median voter?

First, public opinion research—including the reports commissioned by the BPA—parses out what the majority thinks. Following Downs' median voter theorem, one would therefore expect that politicians craft their rhetoric so as to cater to the majority (Wood and Lee

2009). While the Downsian assumptions do not hold in Germany's multi-party system, the main party studied here—Merkel's *CDU*—does consider itself centrist.

Second and related, the provision of public opinion research may prime elected officials to focus their attention toward the median voter. Traditional party politics leads politicians to think about their core clientele. Public opinion research, by contrast, analyzes the entire population. Besides providing quantitative information on majority opinion, this may thus lead politicians' to adopt a more comprehensive view of representation.

- **Hypothesis 2 (*Responsiveness*)**: Exposure to public opinion on a given topic makes elites more likely to endorse majority opinion on the topic

3 Data

3.1 Public opinion research reports

We use an unusual source of evidence to study whether public opinion affects elite rhetoric. In September 2012, Malte Spitz—a member of the German Green party—filed an official inquiry with the BPA. He demanded access to two research reports the BPA had commissioned on behalf of the German government. The office denied the request. The parliamentarian sued the government under the German Freedom of Information Act. He was subsequently granted access to the reports, but was only allowed to make hand copies.

The parliamentarian scanned 125 reports, which comprise over 10,000 pages. An exemplary report page is provided in Figure 1, showcasing their highly detailed nature. The research reports are all addressed directly to chancellor Merkel ("Dear Mrs. Chancellor"), followed by a two-page summary written and signed by the head of the BPA. The purpose of these summaries is to condense the findings so as to give the chancellor and other cabinet members a succinct overview of German public opinion. Most reports include up to 60 pages

of qualitative and quantitative insights. It is here that cabinet members and speechwriters find a rich repertoire to attune speeches to public opinion.

The reports were commissioned by the BPA to leading German public opinion firms before being passed on to the German cabinet. They span the entire legislative term from 2009 to 2013. Importantly, the BPA is an independent German bureaucracy. Contracts with survey firms undergo a tedious tendering process (more below). In our period of study, the BPA commissioned an average of 150 surveys per year. In so doing, the BPA spent an average of two milion Euros per year on its public opinion research (Becker and Hornig 2014). Within the BPA, the public opinion research is overseen by the head of the unit 204 ("public opinion research and evaluation"), Ute Molitor, while the presentation of the findings to the cabinet is overseen by the head of the BPA, Steffen Seibert. Importantly, the Merkel government—unlike the chancellors Kohl and Schröder—specifically decided to commission reports to all major public opinion firms in order to provide a more objective overview of public opinion (Becker and Hornig 2014). That said, Merkel and the cabinet do not directly communicate with the research firms (a detailed discussion of exogeneity is in Section 4.2.1; Becker and Hornig 2014). The reports thus represent a novel lense through which one can assess whether exposure to public opinion affects elite rhetoric.

To validate that the public opinion research reports constitute accurate measures of public opinion, we gained access to Germany's most prominent and longest-running opinion poll: the *Politbarometer*. The *Politbarometer* is commissioned by Germany's public television network (ZDF) and was started in 1977. It consists of regular (about 1.5 per month) surveys that serve to "poll the opinions and attitudes of eligible Germans with regard to current events and issues". During the second Merkel term, the ZDF commissioned 76 surveys with a total of 132,321 respondents. Based on this data, we make two points in Section A.9 on page 15. First, we show that 75 percent of the research reports tackle topics that voters, at the time of the write-up, consider to be a pressing issue. Second, we show that the relative frequency of the topics covered in the *Politbarometer* is highly comparable to the relative

frequency of the topics addressed in the government reports.

To convert the public opinion reports into text data, we followed standard operating procedures in the text analysis literature. Using the R package `tm`, we first removed numbers, punctuation, and white spaces from the reports. We then stemmed all reports by removing stop words and pre- and suffixes. On average, there were 1,683 words per report after stemming.

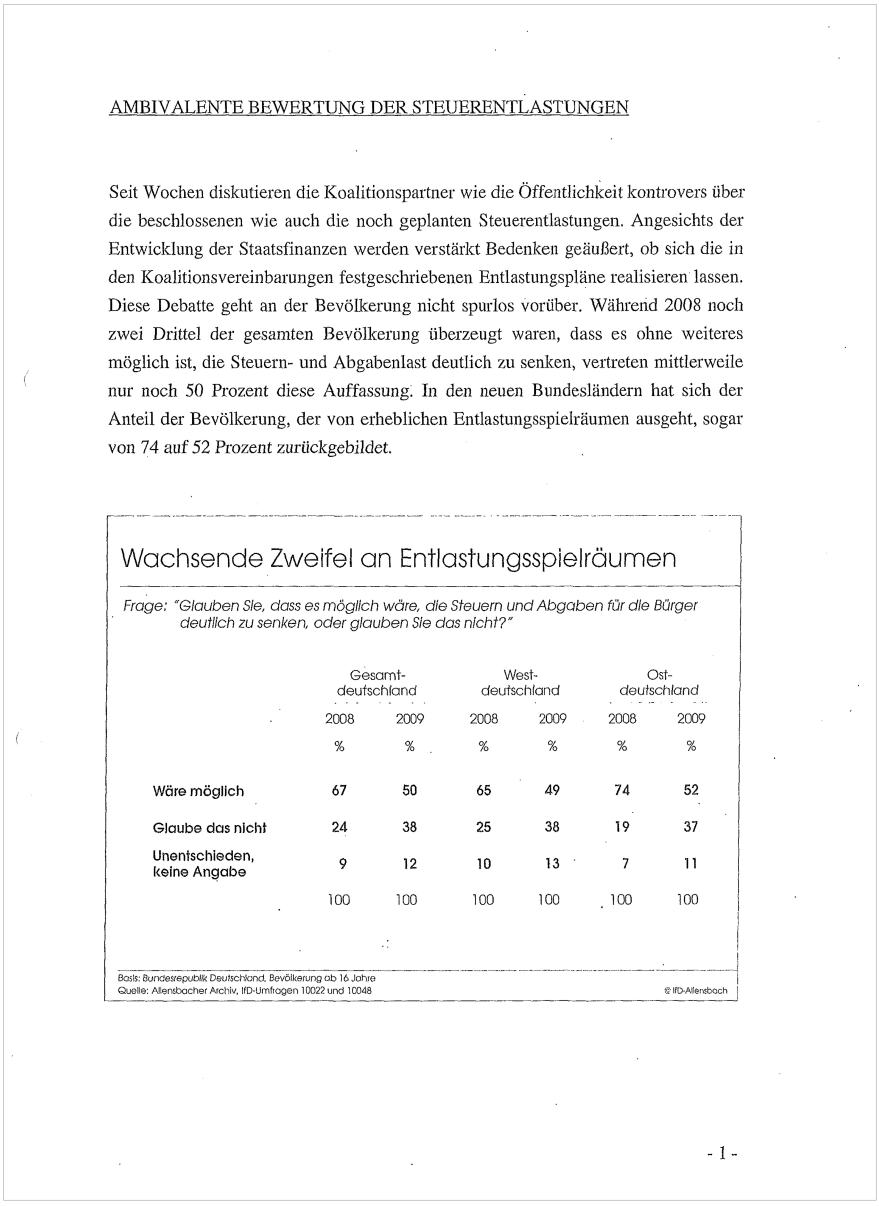
The descriptive statistics of the opinion reports are provided in the left column of Table 1. We asked two independent coders to assign all reports to political topics—a straightforward task given that most reports already include topic headlines. The coders were asked to categorize reports according to the 15 topics used by the German government (*Topic (original)*—more below). To streamline the analysis, we collapsed the 15 topics into 7 categories (*Topic (aggregated)*) that map more intuitively onto German public opinion (Details on the aggregation rule are provided in Section A.1 on page 2.) These include culture (4%), economic policy (35%), education (2%), environmental policy (6%), foreign policy (20%), interior (5%), and social policy (29%).

Table 1 also reports the year and month of the reports' dissemination to cabinet members. Both variables are fairly uniformly distributed. Finally, the table reports which of the seven leading German public opinion firms commissioned the report (*Opinion firm*). There are three key players—Allensbach, Dimap and Emnid—which wrote 72% of all reports.

3.2 Elite rhetoric

We measure elite rhetoric using all speech documents published by the German government from 2005 to 2016—the entire period where Merkel has been in power. These documents include speeches, press releases and articles written by cabinet members. We scraped the documents from the website of the German government (*www.bundesregierung.de*). The resulting data set covers all published speech documents by cabinet members. We applied

Figure 1: Exemplary public opinion report page



AMBIVALENTE BEWERTUNG DER STEUERENTLASTUNGEN

Seit Wochen diskutieren die Koalitionspartner wie die Öffentlichkeit kontrovers über die beschlossenen wie auch die noch geplanten Steuerentlastungen. Angesichts der Entwicklung der Staatsfinanzen werden verstärkt Bedenken geäußert, ob sich die in den Koalitionsvereinbarungen festgeschriebenen Entlastungspläne realisieren lassen. Diese Debatte geht an der Bevölkerung nicht spurlos vorüber. Während 2008 noch zwei Drittel der gesamten Bevölkerung überzeugt waren, dass es ohne weiteres möglich ist, die Steuern- und Abgabenlast deutlich zu senken, vertreten mittlerweile nur noch 50 Prozent diese Auffassung. In den neuen Bundesländern hat sich der Anteil der Bevölkerung, der von erheblichen Entlastungsspielräumen ausgeht, sogar von 74 auf 52 Prozent zurückgebildet.

Wachsende Zweifel an Entlastungsspielräumen

Frage: "Glauben Sie, dass es möglich wäre, die Steuern und Abgaben für die Bürger deutlich zu senken, oder glauben Sie das nicht?"

	Gesamt-deutschland		West-deutschland		Ost-deutschland	
	2008	2009	2008	2009	2008	2009
	%	%	%	%	%	%
Wäre möglich	67	50	65	49	74	52
Glaube das nicht	24	38	25	38	19	37
Unentschieden, keine Angabe	9	12	10	13	7	11
	100	100	100	100	100	100

Basis: Bundesrepublik Deutschland, Bevölkerung ab 16 Jahre
Quelle: Allensbacher Archiv, IfD-Umfragen 10022 und 10048

© IfD-Allensbach

- 1 -

Notes: The Figure displays page 1 of an exemplary research report entitled "Population's view on tax burden and taxation" (*Steuerbelastung und Steuersystem aus Sicht der Bevölkerung*). The English translation is as follows: "For weeks, the coalition partners and the public have been debating controversially over the implemented and planned tax reliefs. In light of the evolution of public finances, concerns have been raised as to whether the austerity plans set out in the coalition agreements can be achieved. This debate has left a mark on the population as well. While in 2008 two-thirds of the entire population were convinced that it would be possible to reduce the burden of taxes and levies, today only 50 percent still hold this view. In the Eastern states, the proportion of the population, which expects considerable relief margins, has even declined from 74 to 52 percent."

Table 1: Descriptive statistics

	Opinion reports		Government speech (manually classified)		Government speech (machine classified)	
	Absolute	*%*	*Absolute*	*%*	*Absolute*	*%*
Total	125	100	9831	100	11973	100
Topic (original)						
Labor / Welfare	32	20.0	509	5.2	-	-
Foreign policy	23	14.4	2168	22.1	-	-
Education / Research	3	1.9	770	7.8	-	-
Agriculture	2	1.2	386	3.9	-	-
Families	11	6.9	437	4.4	-	-
Finances	25	15.6	458	4.7	-	-
Health	4	2.5	345	3.5	-	-
Interior	5	3.1	555	5.6	-	-
Justice	3	1.9	212	2.2	-	-
Culture	6	3.8	1860	18.9	-	-
Environmental policy	7	4.4	537	5.5	-	-
Infrastructure	0	0.0	220	2.2	-	-
Defense	9	5.6	273	2.8	-	-
Economic policy / Energy	30	18.8	785	8.0	-	-
Economic development	0	0.0	316	3.2	-	-
Topic (aggregated)						
Culture	6	3.8	1860	18.9	863	7.2
Economic policy	55	34.4	1463	14.9	1684	14.1
Education	3	1.9	770	7.8	631	5.3
Environmental policy	9	5.6	923	9.4	860	7.2
Foreign policy	32	20.0	2757	28	5406	45.2
Interior	8	5.0	767	7.8	1367	11.4
Social policy	47	29.4	1291	13.1	1162	9.7
Month						
Jan	14	11.2	740	7.5	983	8.2
Feb	6	4.8	774	7.9	904	7.6
Mar	11	8.8	925	9.4	838	7
Apr	10	8.0	797	8.1	1027	8.6
May	9	7.2	791	8.0	1129	9.4
Jun	10	8.0	949	9.7	1213	10.1
Jul	14	11.2	832	8.5	831	6.9
Aug	9	7.2	830	8.4	636	5.3
Sep	12	9.6	907	9.2	1082	9
Oct	7	5.6	794	8.1	1063	8.9
Nov	9	7.2	884	9.0	1319	11
Dec	14	11.2	608	6.2	948	7.9
Year						
2005	0	0.0	8	0.1	0	0
2007	0	0.0	15	0.2	0	0
2008	0	0.0	58	0.6	0	0
2009	22	17.6	38	0.4	828	6.9
2010	29	23.2	221	2.2	2358	19.7
2011	26	20.8	947	9.6	1573	13.1
2012	25	20.0	1535	15.6	742	6.2
2013	23	18.4	1627	16.5	722	6
2014	0	0	1898	19.3	900	7.5
2015	0	0	2017	20.5	1017	8.5
2016	0	0	1467	14.9	1576	13.2
2017	0	0	0	0	2257	18.9
Opinion firm						
Allensbach	34	27.2	–	–	–	–
Dimap	26	20.8	–	–	–	–
Emnid	30	24.0	–	–	–	–
FG	12	9.6	–	–	–	–
GMS	3	2.4	–	–	–	–
Polis	8	6.4	–	–	–	–
TNS	12	9.6	–	–	–	–
Medium						
Article	–	–	5785	58.8	5492	45.9
Press release	–	–	3510	35.7	3219	26.9
Speech	–	–	536	5.5	3262	27.2

Notes: "Absolute" refers to absolute frequencies of categories. "%" refers to relative frequency in percent. For "Opinion firm", "Year" and "Month," the total number of observations is equal to the total number of opinion reports, i.e. N=125. Each opinion report can have up to three topics. Therefore, the topics number N=161. We display the summary statistics for the government speech documents separately for documents that were categorized by the government, and for documents that were automatically classified (see Section A.2 on page 2). We only predict aggregated issue categories for untagged documents.

the same transformation to the text data, i.e., removing superfluous characters and stemming.

The descriptive statistics of the speech data are given in the right column of Table 1. Again, we classify the topic of the speech (*Topic (original)*). In particular, the German government assigned 47% of all documents tags according to the content of the speech. As before, we collapsed these 15 topics into 7 categories. These include foreign policy (28%), culture (19%), economic policy (15%), social policy (13%), environmental policy (9%) as well as interior and education (both 8%). The rest of the documents is published without an explicit topic assignment. To avoid discarding valuable information, we use a machine learning approach to automatically classify the untagged documents. Using a support vector machine that takes the tf-idf document-term matrix of the speech document corpus as the input, we achieve an out-of-sample prediction accuracy of 89% (for details, see Section A.2 on page 2). Table 1 also reports the month and year of the publication. As can be seen, publications increased notably in 2011. The month of the publication, on the other hand, is fairly homogeneous. Finally, Table 1 also indicates the speech documents' type, separating articles (59%), press releases (36%) and speeches (6%).

4 Empirical strategy

Does public opinion affect elite speech? If so, does public opinion merely alter elites' rhetorical agenda or does it also affect the substantive positions they endorse? To address these questions, we use two distinct measurement strategies, which we introduce in turn. Thereafter, we discuss our estimation strategy.

4.1 Measurement

Agenda setting To assess whether the public opinion reports affect elites' rhetorical agenda, we measure whether reports and elite speech become more linguistically similar af-

ter a report has been issued. Formally, this means creating a distance measure $Similarity_{i,j}$ for all public opinion reports i and all speech documents j. This gives $N^{\text{Report}} \times N^{\text{Speech}} = 125 \times (9,831 + 11,973) = 2,725,500$ observations. Our preferred measure of distance is Cosine similarity—a simple and common measure of dissimilarity (e.g., Egesdal, Gill and Rotemberg N.D.). It ranges from 0, documents are entirely dissimilar, to 1, documents are exactly the same.

Responsiveness To assess whether the public opinion reports affect elites' political positions, we measure substantive agreement between reports and elite speech. Our strategy, here, relies on human coding. First, we drew a random sample of 2,000 speech-report pairs (from the set of speech-report pairs that (1) are on the same topic, and (2) were released within 120 days of each other; more below). Second, we asked trained research assistants (who were blinded to the treatment status) to assign each of the pairs a score measuring substantive agreement. The score ranged from -3 (the politician strongly disagrees with public opinion displayed in the report) to 3 (the politician strongly agrees with the public opinion). If elite speech beares no relation to public opinion, the research assistants assigned a score of 0 (more details are provided in Section A.4 on page 4). Importantly, Figure 5 shows that the hand-coded responsiveness measure is unrelated to cosine similarity. This builds trust that both measures tap distinct rhetorical reactions (more below in Section 6).

4.2 Estimation strategy

In order to test whether exposure to public opinion affects elite rhetoric, we exploit the plausibly exogenous timing of the reports' dissemination to cabinet members. Specifically, we construct a regression discontinuity design, which compares elite speech to the public opinion reports right before and after a report is disseminated. Our empirical strategy falls into a broader class of studies that analyze attitudes just before and after exposure to

new information (e.g., Franco, Grimmer and Lim 2017)[1]. We discuss exogeneity and our discontinuity specification in turn.

4.2.1 Exogeneity of report dissemination timing

What determines when reports are sent to the German chancellery? As we outline in the following, there are nine reasons to believe that the precise timing—within a number of days—of the reports' dissemination to the cabinet is exogenous.

First, German law requires that contracts above 5,000 Euros (as is the case for reports of this scope) undergo an official tendering process. Potential contracts must be widely advertised and at least three competitive bids should be received. Since all seven major German public opinion firms are regularly contracted, competition is high. Contractual short-cuts are impossible. The length of the tendering process is thus tough to predict, making it difficult to strategically time the dissemination of findings to the cabinet.

Second, the commissioned reports comprise detailed academic research. They include large representative surveys and detailed focus group discussions. This adds an additional variance to the timing. To assess the influence of data collection effort on the timing of report dissemination in detail, Table A2 on page 6 shows results from a regression of the number of days of data collection on the time it took the BPA to disseminate a report *after* data collection had already concluded. The Table shows that the length of data collection alone explains about 44% of the variation in the timing of the report dissemination.

A third and related check regarding a possibly strategic timing of the dissemination of the reports to cabinet members concerns the timing of the survey evidence contained in the reports and the date of the dissemination. In Figure A.6 on page 8, we plot the number of days between data collection and dissemination to the chancellor's office. The majority of

[1]Franco, Grimmer and Lim (2017) study the effect of Presidential appeals in the US on public approval, constituents' policy preferences, constituents' perceived issues salience and social media activity.

reports are passed on to the chancellor a few days after the last data were collected.[2] This is further evidence that BPA bureaucrats do not strategically time the report dissemination.

Fourth, to quantitatively test for a possible influence of cabinet members on the dissemination timing of the reports, we examine whether reports on more salient topics tend to be disseminated earlier than reports on more peripheral topics. Perhaps, cabinet members try to accelerate the release of reports that cover salient topics. We measure topic salience based on newspaper mentions of a given topic (for details, see Section A.3 on page 4). Reassuringly, Table A2 on page 6 shows that there is no relationship between issue salience and dissemination timing.

Fifth, if the German chancellery does influence the timing of the report dissemination, one would expect them to do so ahead of important elections. To test this hypothesis, we assess whether reports are more likely to be issued right before state elections. Figure A.4 on page 7 shows that this is not the case. If anything, we observe that reports are more likely to be issued *after* elections. The reason might be that research firms are particularly busy ahead of elections and therefore delay the release of the opinion reports to the BPA.

Sixth, in a similar vein the German chancellery could also try to accelerate the dissemination of reports before parliamentary sessions so as to improve their parliamentary speeches. One would therefore expect to see a greater number of reports just prior to days when the German parliament is in session. However, as Figure A.5 on page 7 demonstrates, reports are not more likely to be released in the days preceding parliamentary sessions.

Seventh, if the dissemination date of reports is strategically chosen, this might show up in a non-uniform distribution of release dates across week days or months of the year. Some days and months (notably, Fridays as well as the months of July and August) tend to be

[2]This holds particularly true for the three largest public opinion firms—Dimap, Emnid and Allensbach—which pass on the reports almost instantaneously. As much is shown in Figure A.7 on page 8. Opinion firm with greater timing variance are Polis and TNS, which mostly conduct qualitative interviews and focus group discussions.

less busy in the German political system. However, Table 1 demonstrates a highly regular timing across the year. This further underlines the reading that the BPA tries to provide German political elites with a regular, objective view about German public opinion.

Eight, we can assess whether the covariates of the speech documents are similar right before and after opinion research reports are issued. In Figure A.14 on page 20, we show that opinion report releases induce no meaningful changes across a number of control variables, including government speech topics, timing, length and medium.

Ninth, if the government strategically times the dissemination of the reports, it may do so during times when voters are dissatisfied with the government. In Figure A.13 on page 19 we show that reports are *not* more likely to appear when voters express greater dissatisfaction with the government—as measured in the *Politbarometer*.

4.2.2 Regression discontinuity design

Having made the case that the precise dissemination timing of the public opinion reports to the cabinet is likely exogenous, we leverage this fact by adopting a regression discontinuity design. Conceptually speaking, we restrict our analysis to a short period before and after the dissemination of reports to cabinet members. This means that elite speech that lies around the time of the dissemination has isomorphic potential outcomes. By comparing rhetoric before and after the report dissemination, we can estimate the local average treatment effect (balance tests are presented in Figure A.14 on page 20).

Our main empirical specification is as follows:

$$\mathrm{Y}_{i,j} = \alpha + \tau \mathrm{Exposure}_{i,j} + \beta_1 X_{i,j} + \beta_2 X_{i,j} * \mathrm{Exposure}_{i,j} + \zeta^{'} \mathbf{Z}_{i,j} + \varepsilon_{i,j} \tag{1}$$

Here, $\mathrm{Y}_{i,j}$ is speech document i's cosine similarity or substantive agreement, respectively, with report j. The running variable $X_{i,j}$ is the time in days between the release of the opinion

report and the speech document, which is positive if the speech document is released after the report. $\mathbf{Z}_{i,j}$ is a vector of control variables listed in Table A6 on page 21 and $\varepsilon_{i,j}$ is the error term. We cluster standard errors both by speech document and by opinion report.

Regression discontinuity designs require researchers to choose a bandwidth within which assignment to treatment is plausibly exogenous. Here, we rely on the bandwidth selection method proposed by Calonico, Cattaneo and Titiunik (2014). To obtain the optimal bandwidth, we use a subsample of all report-speech pairs that are released at most 120 days apart and share the same topic. The resulting optimal bandwidth is 22 days.[3] Given that bandwidth calculations are not without criticism, we conduct sensitivity tests in Section 5.2.

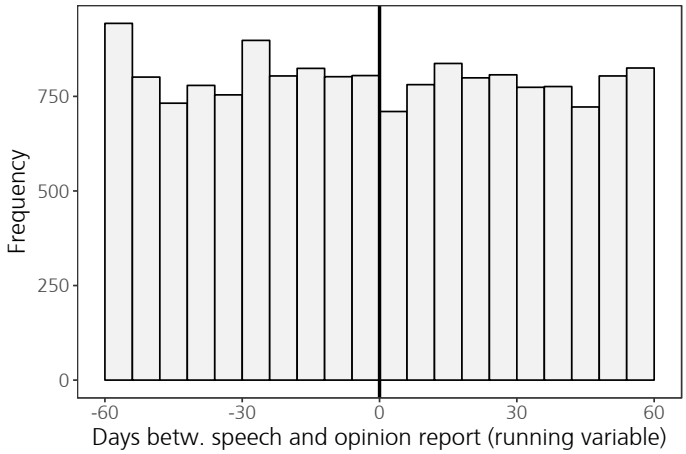
Following Imbens and Lemieux (2008), we estimate our main equation using a local linear regression with the aforementioned optimal bandwidth of 22 days. More specifically, we fit local linear regressions using weights from a triangular kernel to estimate the treatment effect. Imbens and Lemieux (2008) show that the triangular kernel is optimal at the boundary.

A crucial assumption of RD frameworks is that the units of observation have no control over the assignment variable. In our case this would be violated if politicians are able to time the release of a speech conditional on the publication date of an opinion report. To bolster this assumption we follow McCrary (2008) who suggest that researchers estimate whether there is a discontinuity in the number of observations around the threshold. If the density of the assignment variable is not continuous, this may indicate that politicians adjust the timing of speeches to account for the release of opinion reports.

To test the "no jump"-assumption, Figure 2 presents a histogram of the running variable, i.e., we plot the days between the speech document and the opinion report release for a window of 60 days around the release date. The Figure shows that the number of observations

[3]We choose 120 days, rather than the maximum of 4,015 days, because the above-mentioned algorithms otherwise choose bandwidths that are too large (400 days) to make credible inferences. We also confirm our main finding when using all available data—both in a simple OLS as well as when using a multilevel model (see Table A4 on page 19).

Figure 2: Histogram of the running variable



Notes: The Figure plots a histogram of the running variable, i.e., the number of days between speech document release and opinion report release (window of ±60 day).

above and below the cutoff is highly similar. There is thus no evidence of sorting. To formally test this assumption, we use a local polynomial density estimator proposed by Cattaneo, Jansson and Ma (2016) and obtain an insignificant p-value of 0.67.

5 Results

5.1 Agenda setting

We begin by assessing whether exposure to public opinion affects elites' rhetorical agenda. To do so, we assess whether linguistic similarity between public opinion reports and elite speech increases after the dissemination of the reports. Table 2 shows that the dissemination leads to a significant increase in linguistic similarity. Model 1 shows that the dissemination date increases cosine similarity between the reports and elite speeches by 0.014 ($S.E. = 0.007$) on a scale ranging from zero to one. In Model 2, we include all covariates provided in Table 1. The coefficient remains virtually unchanged. This builds trust that our setup is not

compromised by unobserved confounding.

How large are the estimated effect sizes? As Table 2 reports, the effect is akin to a change of 0.14 standard deviations. In Section A.7 on page 14, we report the same models using Jaccard similarity—an alternative measure of linguistic similarity, which captures whether elites incorporate new words from the opinion reports into their speeches. As Table A3 shows, the overlap between government speech and the research reports increases by between 1 and 1.5 percentage points after a report is issued. Overall, we therefore interpret this as evidence that the public opinion reports, indeed, affect elite speech in a substantively meaningful way.

Table 2: RD effects on cosine similarity

	Cosine similarity	
	(1)	(2)
Exposure	0.0137**	0.0128**
	(0.0066)	(0.0057)
Covariates	No	Yes
Observations	5,684	5,684
Mean of DV	0.1263	
SD of DV	0.0976	
Effect size in SD	0.141	0.132

Notes: The Table reports results from a local linear regression around the release of the opinion reports (optimal bandwidth of 22 days; Equation 1). The outcome is the cosine similarity between reports and speeches. The sample is limited to pairs where both speech document and opinion report address the same topic. In Model 2, all covariates reported in Table 1 are included. Standard errors in parentheses are clustered by speech document and by opinion report. *p<0.1; **p<0.05; ***p<0.01

Does the increase in cosine similarity capture agenda setting? While caution is warranted, four reasons undergird this interpretation. First, the analysis focuses exclusively on substantively meaningful words (stop words are excluded). This ensures that the increase in

linguistic similarity is of substantive relevance. Second, to demonstrate this more rigorously, in Section A.6 on page 9 we show that the increase in similarity is driven by substantively meaningful words (following a method proposed by Egesdal, Gill and Rotemberg (N.D.)). For instance, we show that speeches on social policy more frequently rely on the word "social" and refrain from using the word "law" as a result of the dissemination. Third, we do not observe outright plagiarism. Indeed, as Section A.5 on page 8 in the SI shows, elites are careful not to quote verbatim from the reports. The reports are thus not used as a mere rhetorical 'stockpile.' Fourth, the effect sizes are substantively meaningful. If elites merely used the reports to change their wording, one would arguably not expect such noticeable increases in similarity, were it not for a change in rhetorical agendas.

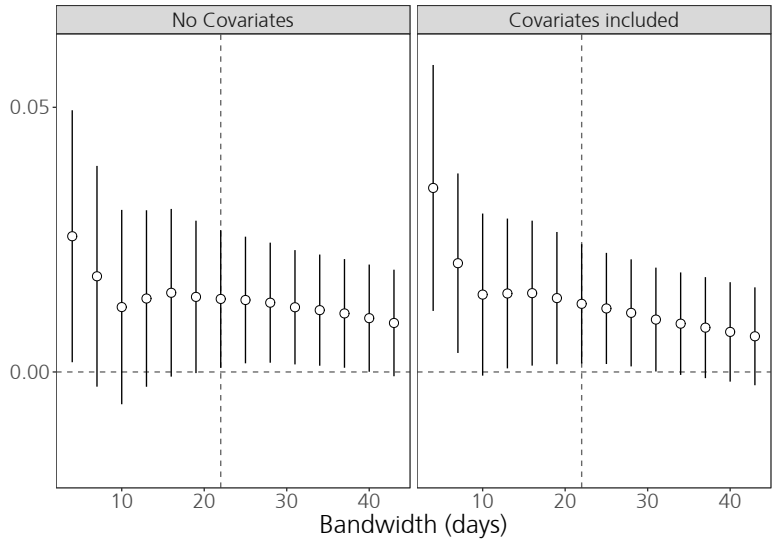
That said, we must reiterate that our analysis matches speeches and reports that address the same broad political topics. We do so because a window of a few days is too small to realistically affect cabinet member's broader speech agenda (we confirm this in Table 3, which we discuss below). As much was relayed to us in qualitative interviews. The chancellor's schedule is set months in advance. A speech scheduled to be delivered in parliament on, say, economic policy cannot be changed to a speech on culture—at least not at such short notice. The public opinion reports can thus only plausibly affect the intensive margin of elites' rhetorical agenda, not the extensive margin. Rhetorical agenda setting thus seemingly takes a more nuanced form: politicians adjust their speeches but within reason. This means, for instance, that a scheduled speech on economic policy, after being exposed to public opinion, changes its focus from growth to taxation. Alternatively, a speech on foreign policy may switch its focus from aid to migration.

5.2 Robustness

Before exploring whether exposure to public opinion also affects elites' substantive positions, we present five robustness tests.

Bandwidth sensitivity In a first step, we let the RD bandwidth vary. Figure 3 provides clear evidence that the dissemination of reports leads to an increase in linguistic similarity—no matter the bandwidth. Even the smallest computationally feasible bandwidth (4 days) shows statistically significant effects. Again, effect sizes are similar when including covariates (right-hand figure). The Figure also shows that the estimate slowly fades out as we increase the bandwidth. This likely demonstrates that the effect of the reports is relatively short-lived, which one would expect given the fast-paced nature of politics.

Figure 3: RD effects on cosine similarity (bandwidth sensitivity)

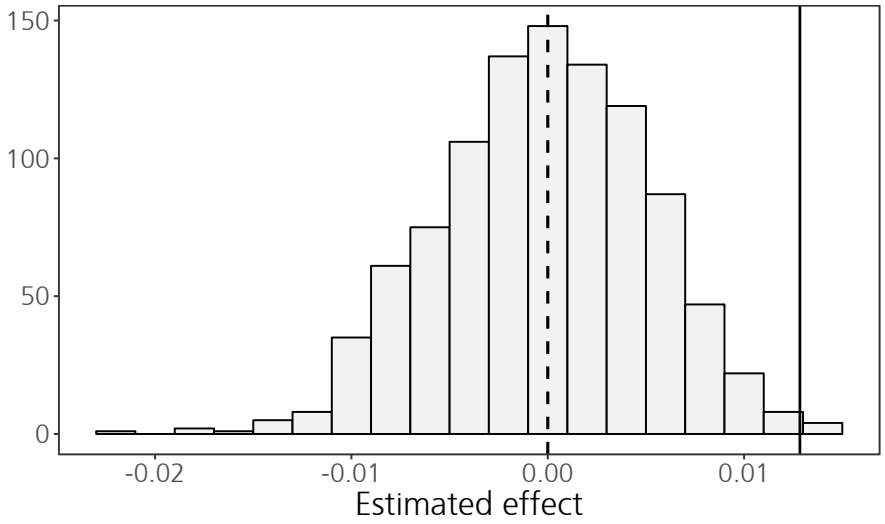


Notes: The Figure plots coefficients and standard errors of RD regression of the cosine similarity outcome on time distance (Equation 1). The x-axis indicates the bandwidth used for the RD estimation. The y-axis plots the estimated effect size and the corresponding standard errors, clustered by speech and report. The vertical dashed line represents the optimal bandwidth (22 days).

Randomization inference In a second step, we implement a version of Fisher's exact test by re-randomizing the "treatment" (i.e., the dissemination date of the public opinion report). To do so, we first generate a list of possible report release dates. Thereafter, we sample new dates from this list with replacement for each report and calculate updated values for the time distance between the speech documents and the reports. We then re-

run the benchmark RD model with the updated time distance, including all covariates, and store the corresponding LATE estimate. We repeat this procedure 1,000 times using the optimal bandwidth of 22 days. In Figure 4, we plot the distributions of the test statistic using the re-randomization procedure. For the optimal bandwidth, we find that only 1.4% of all iterations see a treatment effect that is greater in absolute magnitude than what we see in the actual data (i.e., $P(|\tau^{RD}| > \hat{\tau}^{RD}) = 0.014$).

Figure 4: Randomization inference (cosine similarity)



Notes: The Figure plots the distribution of the LATE from the benchmark RD model when re-randomizing report release dates 1,000 times. The dashed vertical line indicates $\hat{\tau}^{RD} = 0$. The solid vertical line indicates the observed LATE using the original data (see Table 2). The corresponding p-value is 0.014.

Placebo 1: Non-matching topics In a third step, we construct a placebo test that relies on the intuition that one should *not* observe treatment effects when the public opinion reports and the ensuing elite speech address different topics. Given our RD design with a window of a few days, such an extensive margin effect on rhetorical agendas would be all but implausible. As we argued, cabinet members' speeches are planned months in advance. Thus, if one were to observe treatment effects, this would likely imply that unobserved factors such

as general shifts in rhetoric—not exposure to the public opinion reports—are responsible for the observed rhetorical changes. To construct such a test, we use the fact that we know a given opinion report's as well as a given speech document's topic. We then re-estimate our RD model, matching opinion reports and speech documents that do *not* address the same broad topic. For example, this means matching a research report on education to a speech on foreign policy. The results from the placebo test is provided in Table 3. As can be seen, the estimated coefficient is virtually zero. Similarity between elite speech and public opinion reports does *not* increase after the dissemination of reports when the two documents address different topics.

Table 3: RD effects on cosine similarity for unmatched topics (Placebo)

	Cosine similarity	
	(1)	(2)
Exposure	0.0003	0.0015
	(0.0031)	(0.0027)
Covariates	No	Yes
Observations	27,233	27,000
Mean of DV	0.0991	
SD of DV	0.084	
Effect size in SD	0.003	0.017

Notes: The specification follows Table 2; the sample is limited to pairs where speech document and opinion report do *not* address the same topic.

Alternative speech measure In a fourth step, we assess whether exposure to public opinion also affects an alternative measure for elite rhetoric: Speeches delivered in the German parliament. Specifically, we use data provided by Rauh (2015), which covers all speeches given by members of the German government from 2009 to 2013. We focus on parliamen-

tarians that are part of the executive and were thus plausibly exposed to the reports (for details, see Section A.8 on page 15). We use this data to re-estimate our benchmark regression. As Table 4 shows, we see a similar treatment effect for this alternative speech outcome. Interestingly, the estimate is twice as large as our benchmark regression estimate (Table 2). This may be due to the fact that parliamentary speeches receive significant attention in the media, which makes it particularly worthwhile for elites to adjust their rhetorical agenda.

Table 4: RD effects on cosine similarity for parliamentary speeches

	Cosine similarity	
	(1)	(2)
Exposure	0.0487***	0.0331***
	(0.0159)	(0.0117)
Covariates	No	Yes
Observations	2,117	2,117
Mean of DV	0.0866	
SD of DV	0.1016	
Effect size in SD	0.4795	0.3262

Notes: The specification follows Table 2; the sample are the parliamentary speeches.

Table 5: RD effects on cosine similarity for opposition speeches (Placebo 2)

	Cosine similarity	
	(1)	(2)
Exposure	0.0088	0.0142
	(0.0178)	(0.0134)
Covariates	No	Yes
Observations	1,054	1,054
Mean of DV	0.119	
SD of DV	0.1185	
Effect size in SD	0.0747	0.1194

Notes: The specification follows Table 4; the sample is limited to pairs where speech document and opinion report do *not* address the same topic.

Placebo 2: Opposition members In a final step, we use the parliamentary speech data to construct an additional placebo test. Specifically, we assess whether the dissemination of opinion reports affects speeches given by opposition party members. To create a valid counterfactual, we focus on MPs that would plausibly be part of the cabinet—and thus be exposed to the reports—were their party in power (see Section A.8 on page 15 for details). Crucially, since opposition party members are not exposed to the classified reports, we do *not* expect their rhetoric to change. If, however, there are unobserved "rhetorical shocks," one would expect opposition party members to change their rhetoric like members of the executive. The results from our RD estimation are presented in Table 5. As can be seen, we see no significant treatment effect. Opposition party members do not change their rhetoric markedly after reports are issued.

5.3 Responsiveness

We have provided robust evidence that exposure to public opinion leads elites to change their rhetorical agenda along the intensive margin. But is this finding a mere agenda setting effect? Or do elites also adjust their substantive positions? To answer this question, we next turn to the hand-coded set of 2,000 report-speech pairs.

Table 6 reports our benchmark RD model using the substantive agreement outcome. Recall that trained research assistants assigned each pair a score ranging from -3 (the politician strongly disagrees with public opinion displayed in the report) to 3 (the politician strongly agrees with the public opinion). The Table shows that report dissemination is associated with a significant increase in substantive agreement. Model 1 demonstrates that exposure to the reports increases substantive agreement between the reports and elite speeches by 0.2 ($S.E. = 0.08$) on our seven-point scale—a substantively meaningful effect. A precise example of substantive repositioning is Merkel's reaction to a government report, which states "Germans are particularly worried about rising energy prices." Six days later, a press release

by the government states: "She [Merkel] sympathizes with citizens' concerns about rising electricity prices."

And, again, the finding is robust to a variety of sensitivity analyses. First, Model 2 in Table 6 shows that the estimate is highly similar when including all available covariates. Second, Figure A.15 on page 20 demonstrates that the increase is detectable for a wide range of RD bandwidths. Third, Table A5 on page 21 shows that the finding is not detectable when pairing speeches and reports that do *not* share the same topic.[4] Taken together, the robust estimates imply that exposure to public opinion also affects elite's substantive positions. The link from public opinion to elite rhetoric thus seemingly operates through (at least) two channels: agenda setting and substantive responsiveness.

6 Agenda setting vs. substantive repositioning

Can we tell agenda setting and substantive repositioning apart? One concern regarding the analyses presented thus far is that our measures of responsiveness and agenda setting may overlap. Specifically, one may object that if elites adjust their substantive positions, this may also manifest itself in increased linguistic similarity. While this measurement concern does not compromise the causal inferences we draw, it does highlight the difficulty of adjudicating between agenda setting and substantive repositioning. To parse the two rhetorical strategies apart, we provide four pieces of evidence.

First, if substantive realignment goes hand in hand with increased cosine similarity (our measure for agenda setting), one would expect the two measures to be highly correlated. If, by contrast, the substantive agreement measure is orthogonal to cosine similarity, this

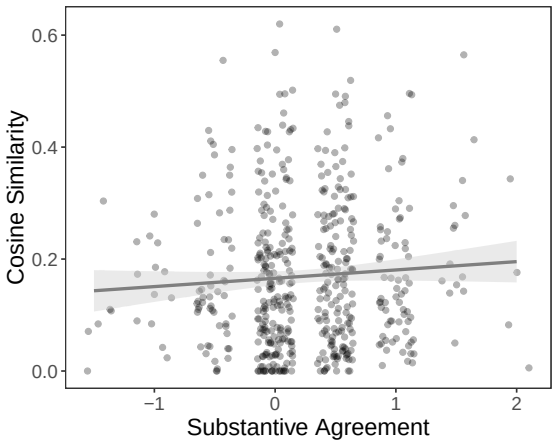
[4]Note that in order to afford this test, we hand-coded an additional random sample of 200 speech-report pairs that do *not* share the same topic. Unfortunately, we are not in a position to repeat the randomization inference robustness test. Doing so would require that we code all 1,316,250 pairs in order to ensure sufficient power around the cutoffs, which is not feasible.

Table 6: RD effects on substantive agreement

	Substantive agreement	
	(1)	(2)
Exposure	0.1952**	0.2048***
	(0.0833)	(0.0710)
Covariates	No	Yes
Observations	214	214
Mean of DV	0.0814	
SD of DV	0.3591	
Effect size in SD	0.5436	0.5703

Notes: The Table reports results from a local linear regression (Equation 1) around the release of the opinion reports (optimal bandwidth of 17 days). The outcome is substantive agreement between reports and speeches. The sample is restricted to pairs where both speech document and opinion report address the same topic. In Model 2, all covariates reported in Table 1 are included. Standard errors in parentheses are clustered by speech document and by opinion report. *p<0.1; **p<0.05; ***p<0.01

Figure 5: Correlation between cosine similarity and substantive agreement



Notes: The plot shows the joint distribution for cosine similarity and substantive agreement, including the line of best fit (OLS). To ease visual interpretation, observations are jittered.

builds trust that the measures capture distinct dimensions of elite rhetoric. Reassuringly, Figure 5 shows no noticeable correlation between the two measures. Increases in substantive agreement (x-axis) are not accompanied by noticeable changes in cosine similarity (y-axis). Put differently, substantive repositioning does not go hand in hand with agenda setting. By the same token, a shift in rhetorical agendas (cosine similarity) does *not* imply that elites also change their substantive positions.

Second, one would expect significant variation across topic salience for the two rhetorical strategies. Salient topics (e.g., migration) tend to polarize the electorate. Here, substantive repositioning may come across as opportunistic and may create reputational costs. Elites may thus be better off putting such topics on their agenda, rather than changing their positions. Doing so allows officials to signal that they take core issues seriously without coming across as a flip-flopper. The case is arguably different for non-salient topics (e.g., culture). Here, agenda setting carries little weight—particularly along the intensive margin. But substantive repositioning may well win over voters without putting one's reputation at risk.

To assess these theoretical considerations, we explore treatment effect heterogeneity by topic salience for the two measures (our salience measure is discussed in Section A.3 on page 4). Table 7 supports our theoretical expectations. For salient topics, we see no substantive repositioning, but increases in agenda setting (cosine similarity). By contrast, among non-salient topics, elites change their substantive positions, but do not engage in noticeable agenda setting. While the differences between these estimates are noisy, they do buttress our intuition that our measures for agenda setting and substantive agreement are distinct and differ in theoretically plausible ways.

Table 7: RD effects (heterogeneity by topic salience)

	Salient Topics		**Non-Salient Topics**	
	Agenda setting	*Substantive agreement*	*Agenda setting*	*Substantive agreement*
Exposure	0.0144*	0.1475	0.0126	0.2123***
	(0.0075)	(0.1257)	(0.0087)	(0.0617)
Covariates	Yes	Yes	Yes	Yes
Observations	2,729	91	2,955	121
Mean of DV	0.1261	0.2198	0.1223	0.1357
SD of DV	0.0973	0.6814	0.0998	0.5348
Effect size in SD	0.149	0.2165	0.126	0.3971

Notes: The specification follows Table 2 and Table 6, respectively; the sample is split as shown.

Third, if public opinion research leads politicians to adjust their rhetorical agendas because of a "cognitive salience"-mechanism (as hypothesized above), treatment effects should arguably increase when elites are repeatedly exposed to the same topic. To explore this conjecture, we proxy for repeated treatments by calculating the dissemination time gap between two reports on the same topic. The median time between reports is 33 days. In Table 8, we re-estimate our benchmark RD regression, splitting the sample along the median. The results

show that the exposure effect is more pronounced when another report on the same issue was released relatively recently. However, the effect is only detectable for the parliamentary speeches. The finding thus provides tentative evidence that agenda setting is, indeed, more likely when elites are exposed to multiple reports within a relatively short time frame.

Table 8: RD effects (heterogeneity by treatment intensity)

	Government releases		**Parliamentary speeches**	
	Repeat treat	*Non-repeat treat*	*Repeat treat*	*Non-repeat treat*
Exposure	0.0114	0.0123	0.0548***	0.0233
	(0.0077)	(0.0102)	(0.0171)	(0.0235)
Covariates	Yes	Yes	Yes	Yes
Observations	1826	2910	1281	1261
Mean of DV	0.1153	0.1303	0.0866	0.0923
SD of DV	0.092	0.1002	0.0966	0.1081
Effect size in SD	0.124	0.122	0.567	0.216

Notes: The specification follows Table 2 and Table 6, respectively; the sample is split as shown.

Last, when politicians engage in agenda setting, one would arguably *not* expect stronger treatment effects when reports contain particularly novel information. After all, we hypothesized that the public opinion research reports lead to agenda setting by a) communicating that the topic is relevant to the electorate and by b) making the topic cognitively salient to the politician. The novelty of the report does not factor into the equation. To proxy the degree to which reports contain new information, we calculate the cosine similarity between a report and the most recent previous report on the same topic. If a report is relatively dissimilar compared to the last report, it likely contains a greater amount of new information. We then re-estimate the benchmark model separately for two subsets, defined by whether the prior report distance measure is above or below the sample median. Reassuringly, Table 9 shows no evidence that report novelty mediates the magnitude of the treatment effect. Put differently, the estimates across the "novel" and "not novel" subsets are not different. We

must caution, however, that this analysis is not conclusive evidence that politicians entirely disregard the novelty of the public opinion reports when setting their rhetorical agendas.[5]

Table 9: RD effects (heterogeneity by novelty of report)

	Government releases		**Parliamentary speeches**	
	Not novel	*Novel*	*Not novel*	*Novel*
Exposure	0.0118	0.0093	0.0628***	0.0435**
	(0.0106)	(0.0084)	(0.0239)	(0.0198)
Covariates	Yes	Yes	Yes	Yes
Observations	1,808	2,208	1,030	1,512
Mean of DV	0.1455	0.1044	0.1027	0.0785
SD of DV	0.1025	0.0873	0.1036	0.1007
Effect size in SD	0.115	0.107	0.606	0.432

Notes: The specification follows Table 2 and Table 6, respectively; the sample is split as shown.

7 Generalizability

Can we characterize whether our results generalize to other contexts? The German electoral system is unique in that it elects MPs both on the basis of plurality vote within electoral districts (majoritarian) as well as on the basis of party lists (proportional representation). Do we observe different treatment effects in the parliamentary speeches for MPs elected via party lists as opposed to MPs elected by majority vote? To answer this question, we code whether MPs were directly elected or not using the aforementioned sample of MPs (see Section A.8 on page 15). We then repeat the analysis shown in Table 4, splitting the sample into directly elected MPs and MPs who entered the parliament via their party's list. Table 10 shows that agenda setting is significantly more pronounced among party list

[5]NB: Changes may, for instance, also occurr when the authors of a report change. Our design is thus best equipped to assess levels (not changes) in public opinion, given that we rely on text data.

candidates. The finding is compatible with our theoretical argument insofar as the opinion reports contain information on the national median voter, which is arguably most relevant to list candidates. Plurality candidates, on the other hand, face constituents whose opinion may differ substantively from the median voter. Therefore, district candidates may be less inclined to incorporate public opinion into their speeches. If taken at face value, the evidence may thus imply that rhetorical agenda setting is most likely to arise in PR systems.

That said, the fact that governments around the world engage in sophisticated public opinion research makes it unlikely that the German case is an outlier (Shapiro and Jacobs 2001). What is more, the observed treatment effects permeate throughout the German political system across different parties and levels of government. Angela Merkel's government is also by no means exceptional in its use of public opinion research. Chancellor Helmut Kohl, for instance, was advised by the Allensbach institute, while chancellor Gerhard Schröder had close connections with the Forsa institute and its then-CEO Manfred Güllner. This makes it less likely that our findings are the product of a highly specific case.

Table 10: RD effects on cosine similarity (heterogeneity by electoral system)

	Cosine similarity	
	Elected in district	Elected through party list
Exposure	0.0291	0.0401**
	(0.0183)	(0.0177)
Covariates	Yes	Yes
Observations	1,437	1,950

Notes: The specification follows Table 2; the sample is split as shown.

8 Discussion

This paper has provided novel text-analytic evidence to assess whether public opinion affects elite rhetoric. Drawing on evidence from Germany, we found that elites change their rhetoric markedly when exposed to public opinion research. Not only does their rhetoric become more similar to the language used in the public opinion reports—a finding that points toward agenda setting along the intensive margin—they also adjust their substantive positions to the public's preferences expressed in the reports. The evidence thus brings clarity to one way through which elites connect with voters: speech.

Before reflecting on the substantive implications of our findings, two words of caution are in order. First, this paper assessed rhetorical agenda setting using cosine similarity, while substantive responsiveness was assessed using human coding. The latter measure, given that it was coded by humans, is rather uncontroversial. Regarding the former, however, a critic might object that cosine similarity does not reflect a true change in agenda, but merely small rhetorical adjustments. We have tried to address this concern by showing that a) similarity increases are driven by substantively meaningful words, and b) that speech writers do not plagiarize from the reports. Still, we must again caution that we merely detect agenda setting along the intensive margin—a result of our local RD design.

Second, our attempt to make a causal argument deserves critical scrutiny. The consistent finding that cosine similarity and substantive agreement increase right after reports are given to cabinet members makes a causal interpretation intuitive. Yet, a skeptic might say that the observed changes are the product of a general shift in rhetoric. While we do not believe that a few days should bring about such drastic changes (indeed, our placebo and permutation tests paint a different picture), the criticism showcases the need to look at rhetoric in a more dynamic setting. Future research could help model such changes with greater clarity, perhaps by benchmarking elite rhetoric to speech in the media. Related, a skeptic might also quibble that our finding is tautological (i.e., elites writing the survey questions and timing

the dissemination of the research). We believe that our qualitative and quantitative evidence rules out this possibility. The pronounced coefficients do underline that elected officials react to public opinion research. At a minimum, our study thus provides descriptive evidence that elites systematically conduct public opinion research and subsequently change their rhetoric.

Having discussed these caveats, we want to briefly reflect on how our research may be expanded on. If German politicians, indeed, adjust their rhetoric to public opinion this bears important insights for the study of representative democracy. In times of increasing polarization, a potential follow-up question is whether citizens perceive such adjustment as deceptive or manipulative. The beginning of the 2000's saw U.S. pundits lament that American politicians were more interested in responding to public opinion than in crafting their own agenda. Such "finger in the wind"-responsiveness was portrayed as signifying a lack of courage and leadership on the part of elected officials (Medvic and Dulio 2004). In the German context, Angela Merkel has been described in *The New Yorker* as "the quiet German"—a politician who silently panders to public opinion (Packer 2014). Similarly, leaked cables show that U.S. diplomats labeled the chancellor "Teflon-Merkel." Merkel's rhetoric, so the story goes, allows her to sidestep political controversy (Waterfield 2010). Sophisticated public opinion data is a double-edged sword. On the one hand, it takes the guesswork out of crafting plebiscitary rhetoric. If taken to the extreme, it can turn political rhetoric into a science that sidesteps truthful dialogue. On the other hand, public opinion research can also work as a means against the sometimes manipulative relationship between speaker and audience created by plebiscitary rhetoric. Such considerations showcase the need to further explore the relationship between elite rhetoric and the public, mapping more fully the ways in which government officials use and interpret public preferences.

References

Becker, Sven and Frank Hornig. 2014. "Kanzler: Regieren nach Zahlen." *Spiegel Online* 37.

Binzer Hobolt, Sara and Robert Klemmensen. 2008. "Government responsiveness and political competition in comparative perspective." *Comparative Political Studies* 41(3):309–337.

Calonico, Sebastian, Matias D. Cattaneo and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82(6):2295–2326.

Canes-Wrone, Brandice. 2006. *Who Leads Whom? Presidents, Policy, and the Public.* Chicago, IL: University of Chicago Press.

Cattaneo, Matias D., Michael Jansson and Xinwei Ma. 2016. Simple Local Regression Distribution Estimators with an Application To Manipulation Testing. Technical report.
URL: *http://www-personal.umich.edu/~cattaneo/papers/Cattaneo-Jansson-Ma_2017_Loc*

Chong, Dennis and James N. Druckman. 2007. "A Theory of Framing and Opinion Formation in Competitive Elite Environments." *Journal of Communication* 57(1):99–118.

Cohen, Jeffrey E. 1999. *Presidential Responsiveness and Public Policy-Making: The Public and the Policies that Presidents Choose.* Ann Arbor, MI: University of Michigan Press.

Egesdal, Michael, Michael Gill and Martin Rotemberg. N.D. "How Federal Reserve Discussions Respond to Increased Transparency." *Working Paper* .
URL: *https://scholar.harvard.edu/files/mrotemberg/files/fed_text_oct2015.pdf*

Eggers, Andrew C. and Arthur Spirling. 2014. "Ministerial Responsiveness in Westminster Systems: Institutional Choices and House of Commons Debate, 1832–1915." *American Journal of Political Science* 58(4):873–887.

Franco, Annie, Justin Grimmer and Chloe Lim. 2017. "The Limited Effect of Presidential Appeals." *Working Paper* .

URL: *https://www.dropbox.com/s/5ym94ggasbff260/public.pdf?dl=0*

Geer, John G. 1996. *From Tea Leaves to Opinion Polls*. New York, NY: Columbia University Press.

Gill, Michael and Arthur Spirling. 2015. "Estimating the Severity of the WikiLeaks US Diplomatic Cables Disclosure." *Political Analysis* 23(2):299–305.

Grimmer, Justin and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21(3):267–297.

Hall, Wynton C. 2002. ""Reflections of Yesterday": George HW Bush's Instrumental Use of Public Opinion Research in Presidential Discourse." *Presidential Studies Quarterly* 32(3):531–547.

Imbens, Guido W. and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142(2):615–635.

Jacobs, Lawrence R. and Robert Y. Shapiro. 2000. *Politicians Don't Pander: Political Manipulation and the Loss of Democratic Responsiveness*. Chicago, IL: University of Chicago Press.

John, Peter and Will Jennings. 2010. "Punctuations and Turning Points in British Politics: The Policy Agenda of the Queen's Speech, 1940–2005." *British Journal of Political Science* 40(3):561–586.

Koch, Jeffrey W. 1998. "Political Rhetoric and Political Persuasion: The Changing Structure of Citizens' Preferences on Health Insurance During Policy Debate." *Public Opinion Quarterly* 62(2):209–229.

Lowe, Will, Kenneth Benoit, Slava Mikhaylov and Michael Laver. 2011. "Scaling policy preferences from coded political texts." *Legislative studies quarterly* 36(1):123–155.

Lucas, Christopher, Richard A. Nielsen, Margaret E. Roberts, Brandon M. Stewart, Alex Storer and Dustin Tingley. 2015. "Computer-Assisted Text Analysis for Comparative Politics." *Political Analysis* 23(2):254–277.

McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* 142(2):698–714.

Medvic, Stephen K. and David A. Dulio. 2004. "The Permanent Campaign in the White House: Evidence from the Clinton Administration." *White House Studies* 4(3):301–318.

Mortensen, Peter Bjerre, Christoffer Green-Pedersen, Gerard Breeman, Laura Chaqués-Bonafont, Will Jennings, Peter John, Anna M Palau and Arco Timmermans. 2011. "Comparing Government Agendas: Executive Speeches in the Netherlands, United Kingdom, and Denmark." *Comparative Political Studies* 44(8):973–1000.

Packer, George. 2014. "The Quiet German." `http://www.newyorker.com/magazine/2014/12/01/quiet-german`. Accessed: 2016-12-11.

Rauh, Christian. 2015. "Communicating Supranational Governance? The Salience of EU Affairs in the German Bundestag, 1991–2013." *European Union Politics* 16(1):116–138.

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson and David G. Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58(4):1064–1082.

Rottinghaus, Brandon. 2006. "Rethinking Presidential Responsiveness: The Public Presidency and Rhetorical Congruency, 1953–2001." *Journal of Politics* 68(3):720–732.

Rottinghaus, Brandon. 2008. "Presidential Leadership on Foreign Policy, Opinion Polling, and the Possible Limits of "Crafted Talk"." *Political Communication* 25(2):138–157.

Shapiro, Robert Y. and Lawrence R. Jacobs. 2001. "Source Material: Presidents and Polling: Politicians, Pandering, and the Study of Democratic Responsiveness." *Presidential Studies Quarterly* 31(1):150–167.

Slapin, Jonathan B and Sven-Oliver Proksch. 2008. "A scaling model for estimating time-series party positions from texts." *American Journal of Political Science* 52(3):705–722.

Vaughn, Justin S. and José D. Villalobos. 2006. "Conceptualizing and Measuring White House Staff Influence on Presidential Rhetoric." *Presidential Studies Quarterly* 36(4):681–688.

Waterfield, Bruno. 2010. "WikiLeaks: Angela Merkel Only Leader Man Enough to 'Rule' Europe." `http://www.telegraph.co.uk/news/worldnews/europe/8168873/WikiLeaks-Angela-Merkel-only-leader-man-enough-to-rule-Europe.html`. Accessed: 2016-12-11.

Weaver, David. 1991. "Issue Salience and Public Opinion: Are there Consequences of Agenda-setting?" *International Journal of Public Opinion Research* 3(1):53–68.

Wlezien, Christopher. 2004. "Patterns of representation: Dynamics of public preferences and policy." *Journal of Politics* 66(1):1–24.

Wood, B. Dan and Han Soo Lee. 2009. "Explaining the President's Issue Based Liberalism: Pandering, Partisanship, or Pragmatism." *Journal of Politics* 71(4):1577–1592.

Wood, Dan B. 2009. *The Myth of Presidential Representation*. Cambridge, U.K.: Cambridge University Press.

A Supporting Information

Contents

A.1 Topic categorization

To distinguish the topics of the government speeches and the opinion reports, we rely on the existing classification of the German government. The German government sorts its own releases into one of 15 topics. We aggregate these 15 topics into 7 broader topics, which we use in all models reported in this paper. In Table A1, we show the aggregation rule.

Table A1: Aggregation of topics

Culture	**Economic policy**	**Education**	**Environmental policy**	**Foreign policy**	**Interior**	**Social policy**
Culture	Finances Infrastructure Economic policy Energy	Education / Research	Agriculture Environmental policy	Foreign policy Defense Economic development	Interior Justice	Labor / Welfare Health Families

Notes: Each column corresponds to one aggregated topic. Column titles are aggregated topics, and column contents are sub-areas that form the aggregated area.

A.2 Classification of untagged speech documents

To assign topics to the untagged elite speech documents (53%), we train an algorithm on the fraction of speech documents ($\approx$ 47%) that was classified by the German government. The input for the classification algorithm is the tf-idf-transformed document-term matrix. We exclude terms with a document frequency lower than 0.01, which leaves us with an input matrix with 3,860 columns. Of the total 10,441 pre-classified speech documents, we set aside 25% as the validation set, and use the remaining 75% to train the classifier. Our outcome variable is the aggregated topic of each speech document, which can take on one of seven values (see Table A1 for an overview).

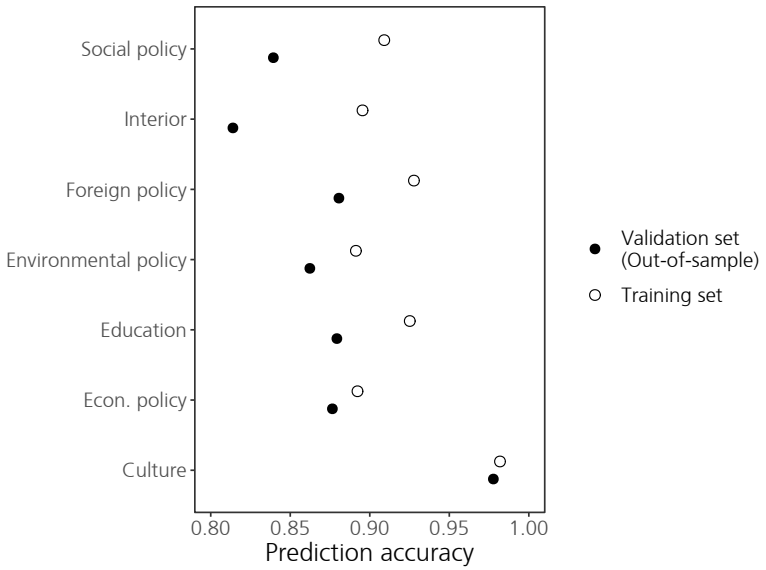
We initially picked two algorithms to predict the topic: The Naive Bayes classifier and the Support Vector Machine (SVM). Here, we focus on the SVM since it performs significantly better than Naive Bayes. To evaluate the performance of the classifier, we choose *prediction accuracy*, i.e., the ratio of correctly classified documents. To avoid over-fitting, we rely on 5-fold cross-validation to find the combination of SVM parameters that maximizes the out-of-sample prediction accuracy.

After training the SVM on the pre-classified speech documents, we evaluate its performance by predicting the speech categories for the 25%-validation set. The SVM is able to correctly classify 92.5% of all speech documents in the training set, and about 88.9% of all

documents in the previously unseen validation set. The high accuracy gives us confidence that the SVM algorithm is able to detect meaningful differences in vocabulary usage and can therefore successfully distinguish between topics. To make sure that the between-topic variation in prediction accuracy is not too high, we examine topic-specific SVM performance in Figure A.1. Here, we see that the prediction accuracy never falls below 80%, regardless of the true topic of the report.[6] This shows that the SVM performs well for each of the topics.

Having confirmed that our prediction method is successful, we then train the SVM on all 10,441 pre-classified speech documents. Subsequently, we predict the speech topics for the remaining 13,238 speech documents that were not tagged by the government. In a final step, we combine the two data sets to obtain the final sample that we use throughout the main analysis.

Figure A.1: SVM prediction accuracy across topics



Notes: The Figure shows the prediction accuracies for the training and the validation sets, based on prediction from a tuned Support Vector Machine model. The dots represent the percentage of correctly predicted document topics, for each of the seven topics that we use in the main analysis. The filled dots represent the prediction accuracy when predicting previously unseen data, i.e., the out-of-sample prediction accuracy, while the hollow dots represent the prediction accuracy for the speech documents that we used to train the algorithm.

[6]Note that we are able to show the out-of-sample prediction accuracy because we split the set of pre-classified documents into a training and a validation set.

A.3 Topic salience

To determine the salience of the seven topics, we proceeded as follows. Using LexisNexis, we collected 7,476 articles released in the leading German weekly newspaper *DER SPIEGEL* between September 2009 and November 2013. We only collected articles that were explicitly tagged as "politics." In addition to the "politics" tag, each article published in the magazine is assigned to up to 51 additional individual topics. This topic categorization is done directly by the authors of the magazine. We manually sorted each topic into one of the seven topics discussed above, wherever possible. As such, individual articles can be assigned to more than one topic. Given article i and topic k, we then generated indicator variables $X_{i,k}$. If topic k is mentioned in article i, $X_{i,k} = 1$, otherwise $X_{i,k} = 0$. To account for articles that discuss more than one topic, we created the new variable $X^*_{i,k}$: $X^*_{i,k} = \frac{X^*_{i,k}}{\sum_{k=1}^{6} X^*_{i,k}}$.

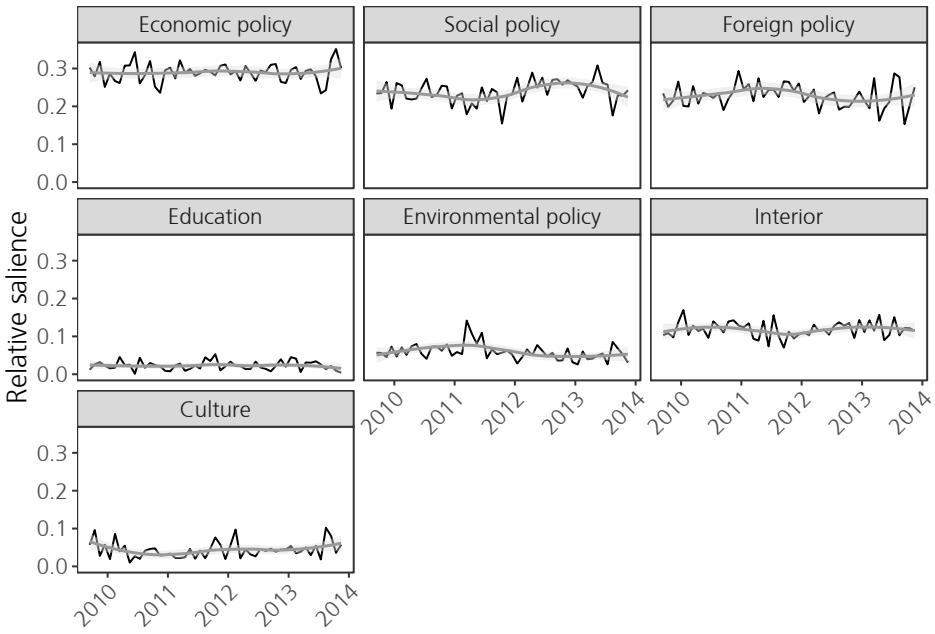
To account for the fact that not all articles are equally relevant, we measured the importance of each article by counting the total number of words in a given article. Let the number of words for article i be l_i. We scaled this measure to range from 0 to 1, such that $l^*_i = F(l_i)$, where F is the CDF of a normal distribution with $\mu = E[l_i]$ and $\sigma^2 = \text{Var}(l_i)$. To measure the salience of issue k within some time frame j, we then calculated: $S_k(j) = \frac{\sum_{i=1}^{N_j} X^*_{i,k}(j) l^*_i(j)}{\sum_{i=1}^{N_j} l^*_i(j)}$

Here, $S_k(j)$ is the salience of topic k within time frame j. As can be seen in the above formula, salience is the relative frequency of articles discussing that topic, adjusted for multiple issues per article and weighted by article length. Our resulting salience measure is provided in Figure A.2. The Figure shows that out of the seven topics, three can be considered salient, namely, economic policy, social policy and foreign policy. At the same time, four topics are significantly less salient. These are education, environmental policy, interior (homeland security), and culture.

A.4 Substantive agreement

To assess whether politicians echo the substantial findings of the opinion reports, we use manual coding. In a first step, we identified all speech–report pairs of interest, i.e., those that share a topic and are released within 120 day of each other. Second, we drew a random sample from this subset of pairs. Third, we gave the sample to two research assistants and asked them to indicate whether the report and the speech document covered the same topic, i.e., whether it is possible to agree or disagree with the opinion research findings in the first

Figure A.2: Issue salience



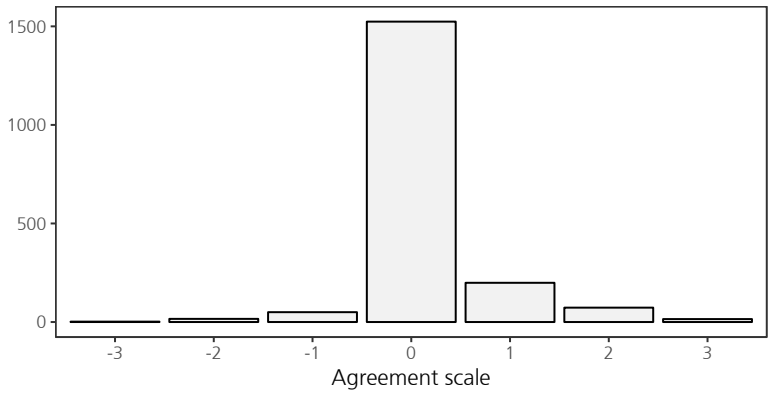
Notes: The Figures plot the relative issue salience per month for the indicated topic (September 2009 to November 2013). The salience measure is constructed from 7,476 articles in *DER SPIEGEL*. Salience denotes the relative number of articles that address an issue. Salience is weighted by the number of issues discussed in the articles (i.e., issue salience decreases when an article discusses multiple issues) and the length of an article (i.e., issue salience increases when more space is devoted to an issue). The grey solid line is a local linear regression smoother.

place. Conditional on this being the case, the coders[7] were then asked to decide whether (i) the politician disagrees with the findings of the report; (ii) the politician agrees with the findings of the; report; (iii) there is no relation between the report and the speech.

Our coding scheme was based on a 7-point scale, ranging from -3 (strong disagreement) to 3 (strong agreement). The midpoint of the scale indicates that the speech document and the opinion report share a topic, but the politician does not explicitly agree or disagree with the findings laid out in the report. Overall, we ended up with a total of 1,705 coded pairs (295 pairs did not match the same topic). In Figure A.3, we plot the resulting distribution of the agreement scores. For the overwhelming majority of documents, we can see that there is either no topical overlap, or there is no substantive discussion of the reports by the politicians – both cases are coded as 0.

[7]The rate of agreement between the two coders was relatively high: For about 80% of all observations, the two coders completely agree. For more than 90 percent of all observations, the coders assign values that are within one unit of each other.

Figure A.3: Distribution of substantive agreement



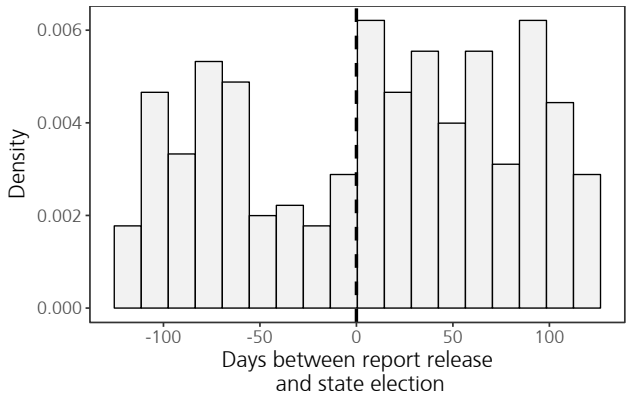
Notes: The plot shows the distribution of the substantive agreement outcome variable, as described in Section A.4 on page 4. Higher values indicate that government speech echoes findings from the opinion reports.

Table A2: Determinants of report timing

	Days between data collection and report release						
Salience (SD)	−3.049			2.145			2.486
	(2.197)			(4.272)			(3.325)
Report length (SD)		3.662*			4.161**		1.748
		(2.064)			(1.977)		(1.376)
Length of collection			1.540***			1.392***	1.139***
			(0.160)			(0.166)	(0.161)
Economic policy				−11.614**	−7.340*	−4.600	−7.925**
				(4.731)	(4.120)	(3.264)	(3.720)
Foreign policy				−18.045***	−15.671***	−7.020*	−8.601**
				(4.348)	(4.776)	(3.952)	(3.553)
Culture				8.682	18.823**	14.609**	12.217
				(11.973)	(9.171)	(7.274)	(9.426)
Education				21.006	15.145	7.711	13.633
				(14.317)	(12.743)	(10.164)	(11.006)
Interior				−3.393	−4.716	−3.738	−0.661
				(8.048)	(8.406)	(6.711)	(6.190)
Environmental policy				−12.060	−10.501	−4.366	−2.370
				(8.193)	(7.661)	(6.132)	(6.520)
N	82	120	120	82	120	120	82
R-squared	0.024	0.026	0.440	0.313	0.207	0.495	0.611

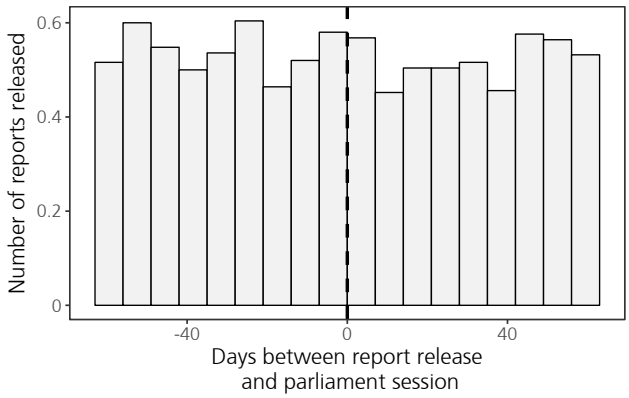
Notes: The dependent variable is the number of days between the last day of data collection and the dissemination date of the report. Topic salience is measured based on a given topic's media salience at the time of the first day of data collection (see Section A.3 on page 4). The length of the report is measured using the number of words after stemming (the variable is standardized). The second set of variables are dummies for the six topic areas, with social policy as the reference category. Standard errors are given in parentheses. ***p < .01; **p < .05; *p < .1.

Figure A.4: Opinion report release timing and state elections



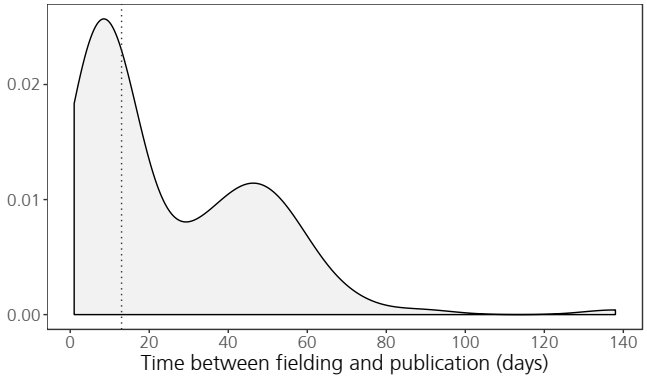
Notes: The Figure plots a histogram of the days between the opinion report release and the date of a state parliament election. Only observations within a window of ±120 days are included. Negative values on the x-axis indicate that a report was released prior to a state election. The vertical dashed line marks the day of the state election. The total number of state elections between 2009 and 2013 is 20. Each bin corresponds to a two-week interval.

Figure A.5: Opinion report release timing and parliament session days



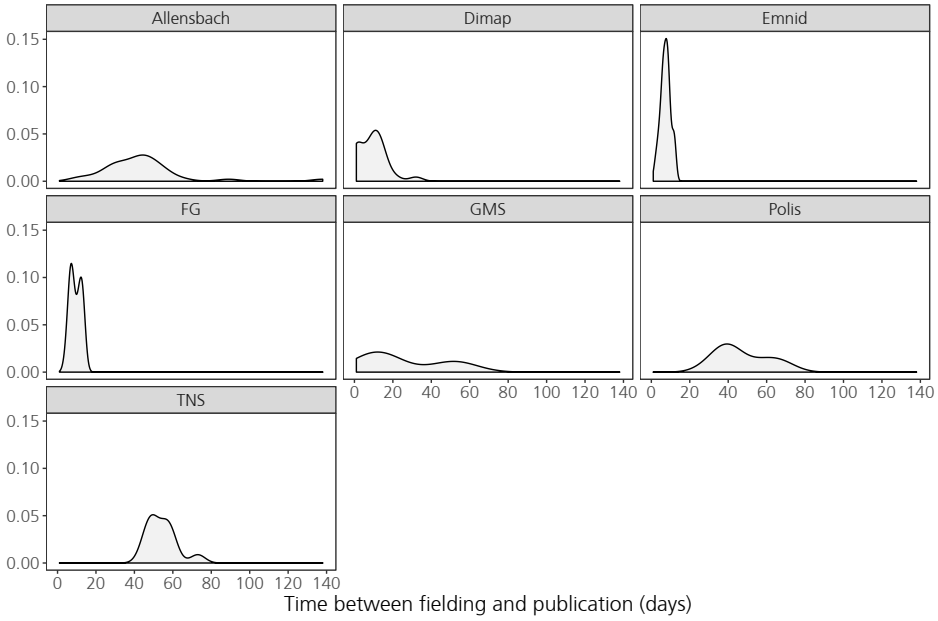
Notes: The Figure plots a histogram of the days between the opinion report release and the days when the parliament is in session. Only observations within a window of ±60 days are included. Negative values on the x-axis indicate that a report was released prior to a day on which the parliament was in session. The vertical dashed line marks the day when the parliament was in session. We consider all session day–report releases pairs. The height of each bin has been divided by the total number of session days. Each bin corresponds to a one-week interval.

Figure A.6: Time between data collection and report dissemination



Notes: The Figure plots the distribution of time between the last day of data collection for a given report and the date when the report was disseminated to cabinet members. The dotted vertical line is the median number of days (=13 days).

Figure A.7: Time between data collection and release date (by opinion firm)



Notes: The Figures plot the distribution of time between the last day of data collection for a given report and the date when the report was disseminated to cabinet members for each survey firm.

A.5 Verbatim quotations

In this section, we assess whether politicians use direct quotes from opinion reports in their speeches. To do so, we use a metric that is commonly used in plagiarism detection: The

longest common subsequence (LCS) distance. The LCS is a method that searches for the longest sequence of words that appears *both* in an opinion report and a speech document. As in the main analyses of the paper, we only consider the subset of opinion report–speech pairs that are at most 120 days apart (i.e., a window within which politicians could plausibly have accessed the opinion reports.)

This leaves us with a list of LCSs for all speech report pairs where (1) report j was released prior to speech document i and (2) speech document i was released not more than 120 days after report j. For each speech–report pair that fulfills these two conditions, we obtain the LCS. Before calculating the LCS, we remove all numbers, special characters and punctuation. However, we do not stem the documents. After executing the procedure described above, our first step is to discard all pairs where the LCS is 10 or smaller. Of the speech report pairs that satisfy conditions (1) and (2), only about 7% have LCSs with more than 10 characters.

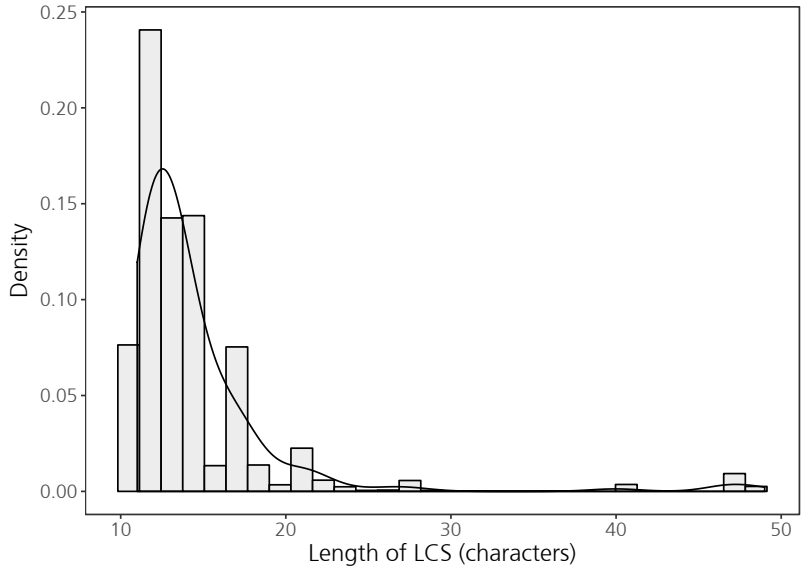
In Figure A.8, we present the distribution of the LCS metric for the set of speech–report pairs where the LCS is longer than 10 characters. As the plot shows, there are very few speech-report pairs for which the longest common substring has more than 20 characters. Given that German words tend to be relatively long, an LCS metric of 20 or lower usually corresponds to a sequence of, at most, three words. Such short sequences are unlikely to be direct quotes from the opinion reports. The distribution of the LCS metric therefore indicates that there appears to be little verbatim quoting from the opinion reports.

A more detailed look at the longest common subsequences supports the observation that verbatim quotes are all but absent. Almost all of the sequences that consist of more than 20 characters are versions of either "Presse und Informationsamt der Bundesregierung" (German Federal Press Office) or "Europäischen Union" (EU). There is hence little reason to believe that politicians quote verbatim from the opinion research reports.

A.6 Word contribution

To better understand the increase in cosine similarity, we assess which words drive the observed increase in similarity. To do so, we analyze how the growth in similarity between speech documents and opinion reports can be decomposed into the contribution of individual words, following a method proposed by Egesdal, Gill and Rotemberg (N.D.). Let the set of all words (after removal of sparse words) be K. We now define the "gap" in usage of word $k \in K$ between speech document i and opinion report j as follows:

Figure A.8: Distribution of the LCS metric



Notes: The Figure plots the distribution of the LCS metric for speech–report pairs where the LCS consists of more than 10 characters. The LCS is the longest sequence of consecutive characters that appears both in a speech documents and in an opinion report. A higher LCS metric indicates that the longest common subsequence could be a direct quote. The solid lines is a kernel density estimate of the distribution.

$$w_{i,j}^k = \frac{n_i^k n_j^k}{x_i \cdot x_j} - \frac{(n_i^k)^2}{\|x_i\|^2}$$

where n_i^k and n_j^k correspond to the number of times word t is used in speech document i and report j, respectively. In addition, $x_i \cdot x_j$ represents the dot product of the term-document vectors[8] of speech i and report j. Finally, $\|x_i\|^2$ is the Euclidean norm of x_i.

The interpretation of $w_{i,j}^k$ is straightforward: If $w_{i,j}^k$ is greater than zero, the word k is relatively over–represented in speech i compared to report j. If $w_{i,j}^k$ is smaller than zero, word k is under–represented in speech i. Intuitively, two documents are more similar when the same words are used with approximately the same frequency in both documents. Similarity increases when the degree of over– or under–representation of certain words becomes smaller over time.

To compute how much each word contributes to the overall increase in cosine similarity, we again use a regression discontinuity approach. The RD method can tell us whether there is a discontinuous decrease in the word usage gap w^k shortly after the release of each opinion

[8]The term–document vector is a vector of length $|K|$, where the k^{th} element corresponds to the number of times word k is used in that document. Therefore, the k^{th} element of x_i is n_i^k.

report.[9] More formally,

$$\tau_{RD}^{k} = E[w_{i,j}^{k}(1)|t_{i,j} = 0] - E[w_{i,j}^{k}(0)|t_{i,j} = 0]$$

Here, $t_{i,j}$ is the time between the release of the speech document and the release of the opinion report. As before, $w_{i,j}^{k}(1)$ is the word usage gap for speech–report pairs where the speech was released after the opinion report, and $w_{i,j}^{k}(0)$ is the word usage gap for pairs for which the report was released after the speech document.

Over time, speeches and reports will only become more similar when the frequency of word usage becomes more equal across documents. This is the case when $E[w_{i,j}^{k}(1)] > E[w_{i,j}^{k}(0)]$ for $E[w_{i,j}^{k}(0)] < 0$ (underrepresented terms are employed more frequently) and $E[w_{i,j}^{k}(1)] < E[w_{i,j}^{k}(0)]$ for $E[w_{i,j}^{k}(0)] > 0$ (overrepresented terms are used less often). Therefore, we can define $\Delta w^{k} = |E[w_{i,j}^{k}(0)]| - |E[w_{i,j}^{k}(1)]|$. If $\Delta w^{k} > 0$, term k will contribute to greater linguistic similarity. To compute the influence of each word, we calculate $\Delta w^{k} =$ for all terms $k \in K$. We first divide all speech-report pairs into our seven main topics. For each topic we then compute Δw^{k} for all terms.

Figure A.9 presents our results. We rank each term by its contribution to the overall change in cosine similarity: The terms at the top of each respective graph have the largest Δw^{k}. Those are the words for which the "gap" in usage between speeches and reports has changed the most after the release of the report. We also distinguish whether cosine similarity increases because a word has been used more frequently (i.e., $E[w_{i,j}^{k}(1)] > E[w_{i,j}^{k}(0)]$, indicated by a "+") or less frequently ($E[w_{i,j}^{k}(1)] < E[w_{i,j}^{k}(0)]$, indicated by a "-").

Importantly, the Figure shows that the rhetorical adjustment stems from substantively salient words. For instance, speeches on economic policy substitute words such as "energy" for words such a "child" after reports have been issued. Speeches on social policy more frequently rely on the word "social" and refrain from using the word "law" as a result of the dissemination. Overall, the analysis thus shows that increases in similarity are not mere statistical artifacts or driven by irrelevant words or phrases. On the contrary, the reports arguably affect the rhetorical agenda of politicians.

[9] We use the optimal bandwidth to estimate the RD. However, there are too few observations around the optimal bandwidth for three topics (interior, education and environmental policy). For these topics, we consider a discontinuous jump at twice the optimal bandwidth (56 days).

Figure A.9: Word-level contributions to increased cosine similarity [Part 1]

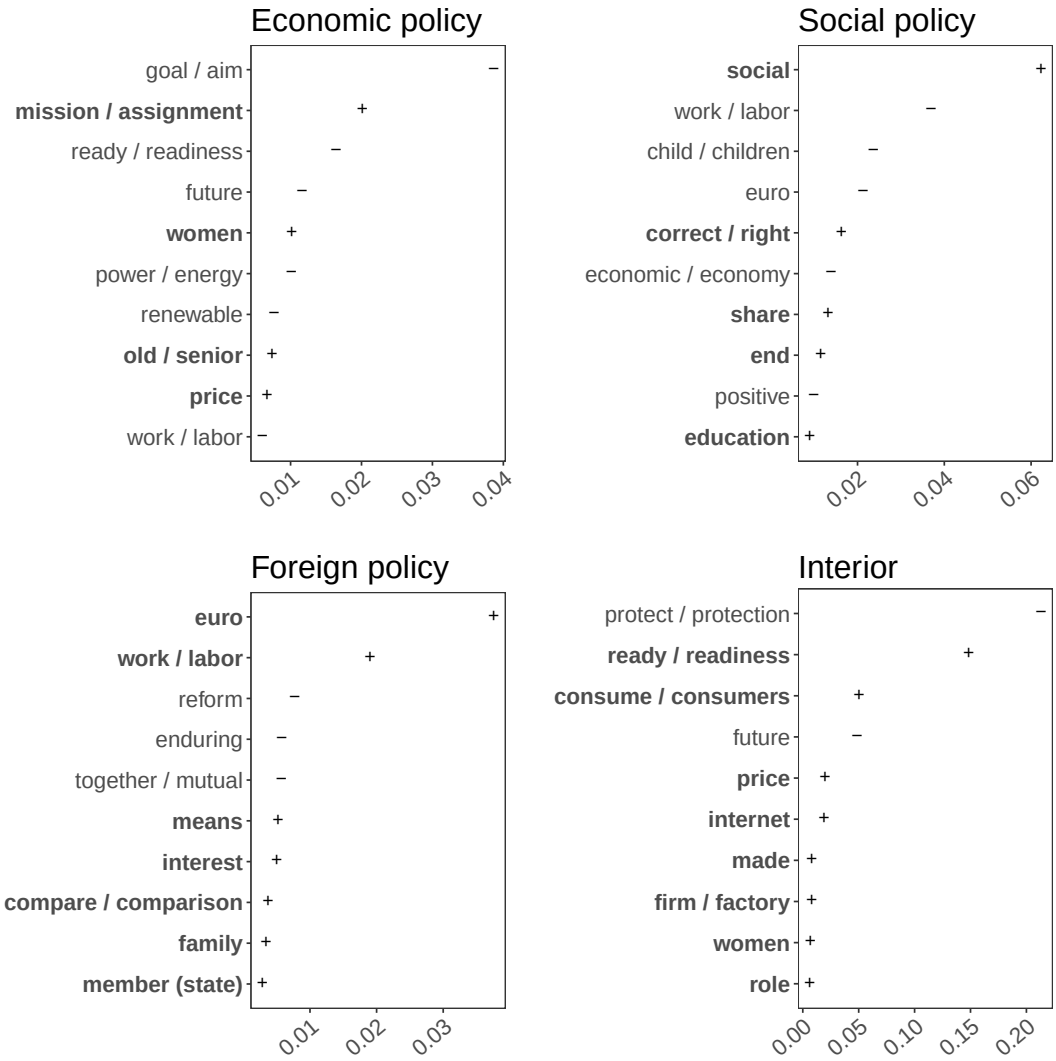
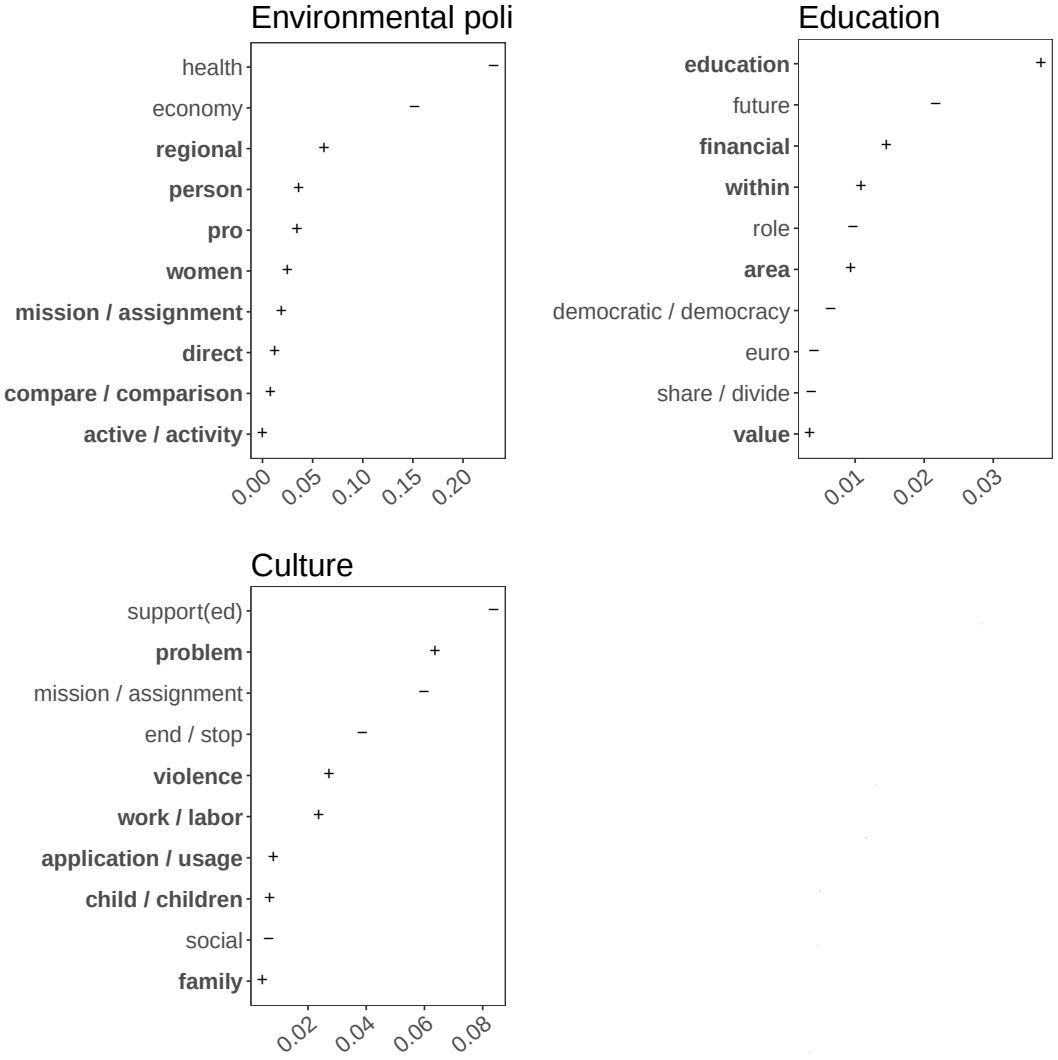


Figure A.12: Word-level contributions to increased cosine similarity [continued]



Notes: The Figure plots increases in cosine similarity after reports have been released, plotting the contributions of individual words. "+" indicates increased use of a given word, while "-" indicates decreased use of a given word. The measure of word contribution is based on a discontinuous jump in word usage before and after the release of the opinion report. The x-axis indicates decreases in word usage dissimilarity. Greater values indicate that usage of a specific terms becomes more similar across speech–report pairs. Only speech-report pairs that share the same topic are considered. The terms have been translated from German. When a word has multiple meanings, the two closest translations are given. The analysis uses the sample of manually coded speeches to minimize noise.

A.7 Jaccard Similarity

To make our findings more interpretable, we present an alternative measure of linguistic similarity: Jaccard similarity. It measures whether the government incorporates new words from the opinion reports in their releases. Let A_i be the set of words used in speech document i, and A_j be the set of words used in opinion report j. Jaccard similarity is defined as $\frac{|A_i \cap A_j|}{|A_i \cup A_j|}$. Crucially, Jaccard similarity is based on whether a given word is used at all, not how many times it is used. Therefore, it is less suited to pick up on general changes in the use of language. However, it can give a sense of whether exposure to public opinion makes politicians use terms that they had not been using prior to being exposed to the opinion reports. In the first two columns of Table A3, we repeat our main analysis with Jaccard similarity as the dependent variable. We find that exposure to public opinion increases Jaccard similarity by between 0.01 and 0.015 units.

Table A3: RD effects on Jaccard similarity and number of words

	Jaccard similarity		*Individual words*	
	(1)	(2)	(3)	(4)
Exposure	0.0153*	0.0102**	3.9612	1.9998***
	(0.0084)	(0.0036)	(2.4921)	(0.6408)
Covariates	No	Yes	No	Yes
Observations	1,182	1,182	1,182	1,182

Notes: The Table reports results from a local linear regression around the release of the opinion reports (optimal bandwidth of 22 days; Equation 1). The outcome is the Jaccard similarity (models 1 and 2) or the absolute overlap (models 3 and 4) between reports and speeches . The sample is limited to pairs where both speech document and opinion report address the same topic. In models 2 and 4, all covariates reported in Table 1 are included. Standard errors in parentheses are clustered by speech document and by opinion report. *p<0.1; **p<0.05; ***p<0.01

In addition, Table A3 also assesses the absolute overlap between speech documents and opinion reports. We define this as the numerator of the Jaccard similarity measure, i.e. $|A_i \cup A_j|$. The results show that exposure to public opinion induces politicians to use between two and four new words taken directly from opinion reports. These are words that were previously not part of the elite speech documents. We stress, however, that (1) this only measures whether politicians use the word, not how frequently they use the words and (2) these are words that were not used previously. Therefore, this number likely underestimates the total number of times a term from the opinion report is used in government speech.

A.8 Parliamentary speech data

To corroborate the finding that exposure to public opinion affects elite speech, we also analyze speeches given in the German parliament. We use data provided by Rauh (2015), which covers all speeches given from 2009 to 2013. In the following, we discuss (1) the sample composition and (2) the method we use to automatically assign topics to the speeches.

Sample composition Given that the public opinion reports are only passed on to cabinet members, we only look at speeches given by the chancellor, government ministers and state secretaries. We also examine the effect on speeches delivered by members of the opposition party (Placebo 2; see Section 5.2). To create a valid counterfactual group, we select high-ranking opposition party members—politicians that would likely be members of the executive were their respective parties in power. Here, we include politicians who were either (1) head of their parties (*Parteivorsitzende*), (2) head of a parliamentary group (*Fraktionsvorsitzende*) or (3) head of a parliamentary committee (*Ausschussvorsitzende*).

Predicting parliamentary speech topics The parliamentary speeches are not assigned specific topics. Since our analysis relies on the comparison of documents that address the same topic, we therefore must assign the speeches to one of the seven issue categories laid out in Table A1. Given that there are 58,361 parliamentary speeches, we use a machine learning approach that mirrors the prediction task described in Section A.2. We train a topic prediction algorithm on the government speech documents that were already tagged by the government. To train the algorithm, we use the tf-idf-transformed document-term matrix of the pre-classified government releases as the input.In a final step, we rely on the algorithm to predict the (unknown) topics of the parliamentary speeches (for more details, see Section A.2 on page 2).

A.9 Validating Government Public Opinion Research

To validate the government public opinion research reports, we use a prominent German opinion poll, the *Politbarometer*. The *Politbarometer* is commissioned by Germany's public television network (ZDF) and was started in 1977. It consists of regular (about 1.5 per month) surveys that serve to "poll the opinions and attitudes of eligible Germans with regard to current events and issues as well as to political parties and individual politicians." During the second Merkel term, the ZDF commissioned 76 surveys with a total of 132,321 respondents, which we download and aggregate.

First, we compare the topic choice in the government reports with the reported issue salience in the *Politbarometer* surveys. This allows us to determine whether the BPA selects topics that track current public sentiments. To measure issue salience among the public, we use a *Politbarometer* item which asks respondents to state what they perceive to be the most pressing issues in German politics. We sort respondent replies into our seven issue categories (see Table A1), and compute the perceived salience of each issue over time. We then assign this survey-based salience measure to each government opinion report based on the issue that the report covers.

In Figure A.10, we present the results: The figure shows that opinion reports overwhelmingly cover high-salience issues. On average, the likelihood that an opinion report will cover either the most or second most salient topic is about 77%. Conversely, only about one fourth of reports cover issues that are not among the two most salient issues at the time of data collection. Figure A.11 shows issue salience over time, separately for each topic. In each panel, we also show the frequency of report releases over time, indicated by the vertical lines at the bottom of each panel. As in Figure A.10, the data shows that the government mainly conducts opinion research on the two most salient issues, economic policy and social policy.

We now examine whether the government research reports address topics that follow the public's interest. We aim to examine whether governmental opinion research covers similar topics as the Politbarometer. As described in Section 3.1, we aggregate opinion report topics into seven issues categories as shown in Table A1. For the Politbarometer, we create a similar measure. Since the within-survey variation in issue coverage is large, we code over 1,000 Politbarometer items and then aggregate the hand-coded survey items, creating a fractional measures of issue coverage for each Politbarometer wave. We normalize the relative frequency of each issue in a given survey by the relative frequency of questions that can be assigned to one of those seven issue categories. We then aggregate the relative frequencies of the seven issues over all 76 Politbarometer surveys.

In Figure A.12, we compare the relative frequency of issues categories between the Politbarometer surveys and the government opinion research. We find that issue choices are similar: Economic policy, social policy and foreign policy dominate the opinion reports and the Politbarometer surveys. Overall, the distribution of issue coverage is highly similar across the government research reports and the Politbarometer surveys.

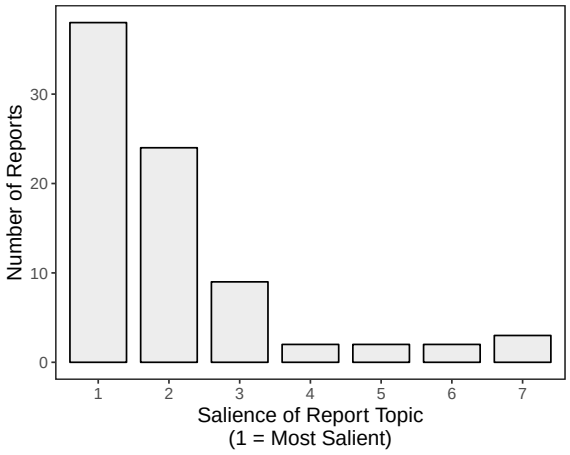
Third, we use the Politbarometer surveys to examine whether the government commissions opinion reports as a reaction to their own popularity. The Politbarometer surveys ask respondents to state their satisfaction with the work of the federal government on a scale ranging from -5 (very dissatisfied) to 5 (very satisfied). We average survey responses to

estimate a simple model of report timing as a function of weekly average public satisfaction with government performance. Since the Politbarometer is not a weekly survey, we use the most recent prior Politbarometer satisfaction measure for weeks in which no Politbarometer survey was conducted.

In Figure A.13, we present results from four OLS models, where we present specification with different combinations of covariates and fixed effects, as well as lags to allow for a delay in opinion report commissioning. We cannot reject the null hypothesis that there is no relationship between satisfaction with the government and the commissioning of governmental opinion research. These results show that governmental opinion research cannot be considered a reaction to declining satisfaction with the government.

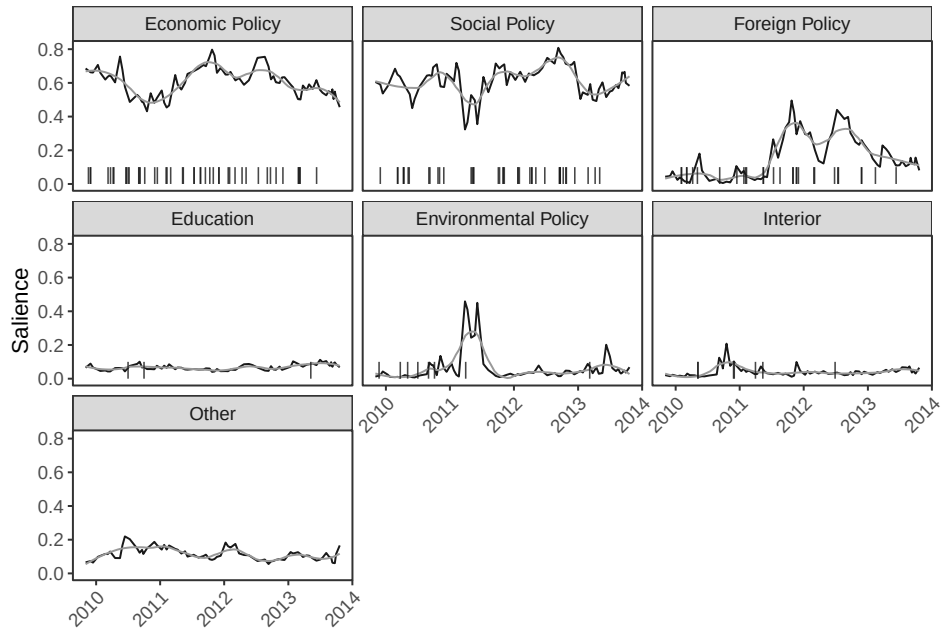
Concluding our analysis, we find that the government broadly commissions reports that cover topics that the public considers to be salient. There is considerable overlap between the issues covered in the Politbarometer and the government opinion research. However, we find little support for the hypothesis that governmental opinion research can be considered a reaction to low levels of popularity.

Figure A.10: Number of reports conditional on salience rank of report topic



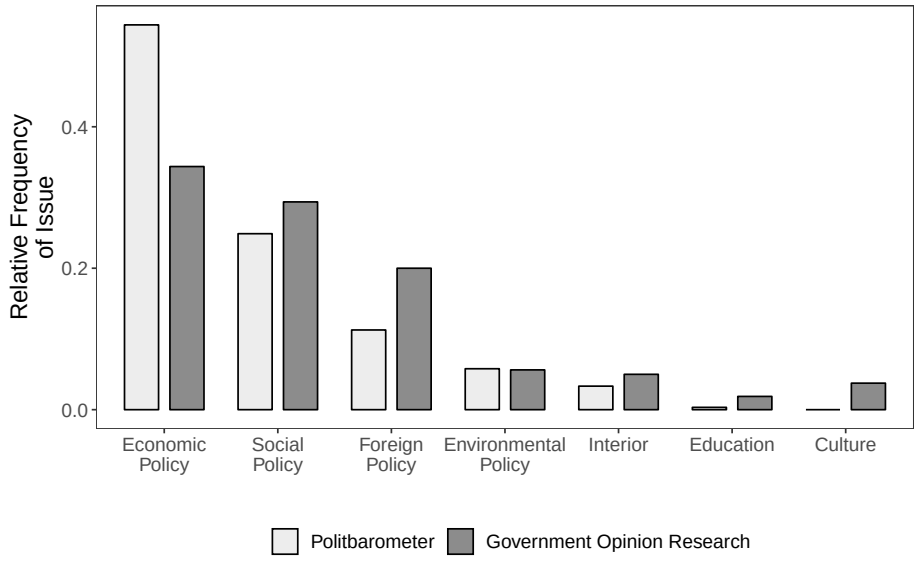
Notes: The Figures shows the number of government reports conditional on the salience of the report topic at the time of data collection ('1' is most salient; salience is based on *Politbarometer* data). The y-axis is the absolute number of reports released.

Figure A.11: Report dates and issue salience



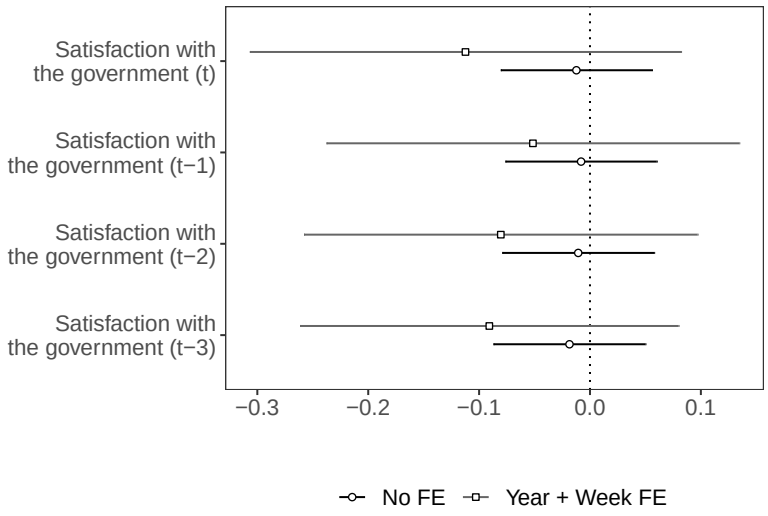
Notes: The panels show issue salience over time based Politbarometer responses. The x-axis is the date, the y-axis indicates issue salience. We plot the raw salience measures (faint grey lines) as well as a LOESS smoother of over-time changes in salience (light and dark blue lines). The vertical lines at the bottom of each panel indicate the start dates of the data collection for the governmental opinion reports.

Figure A.12: Relative frequency of issues covered in *Politbarometer* and government reports



Notes: The Figure shows the relative frequency of issues categories covered in the Politbarometer (light grey bars) and the government opinion reports (dark grey bars).

Figure A.13: Report dates and government satisfaction



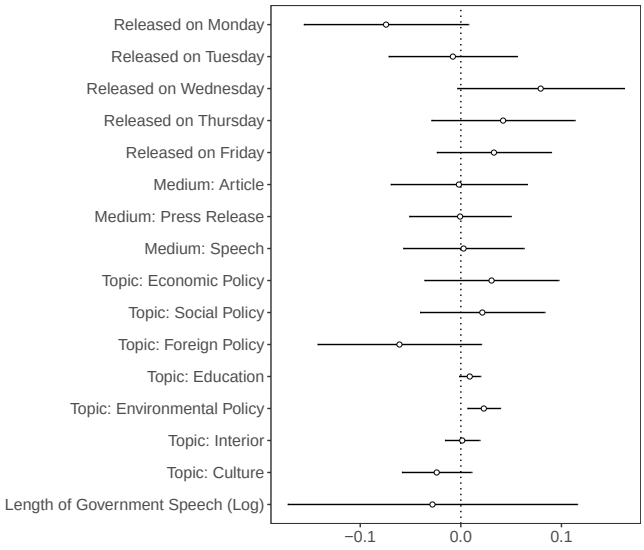
Notes: The Figure shows the effect of a one-standard deviation increase in satisfaction with the government (at different points in time; measured via the *Politbarometer*) on the probability that the government will commission an opinion report. The unit of observation is the week; the outcome is whether a report is commissioned or not.

Table A4: Effect of exposure (across time) on cosine similarity

	Cosine similarity	
	OLS	Multilevel Model
Time since exposure (months)	-0.0006**	-0.0008*
	(0.0003)	(0.0005)
Covariates	Yes	Yes
Observations	109,140	109,140
Mean of DV	0.1212	
SD of DV	0.0963	
Effect size in SD	0.0062	0.0083

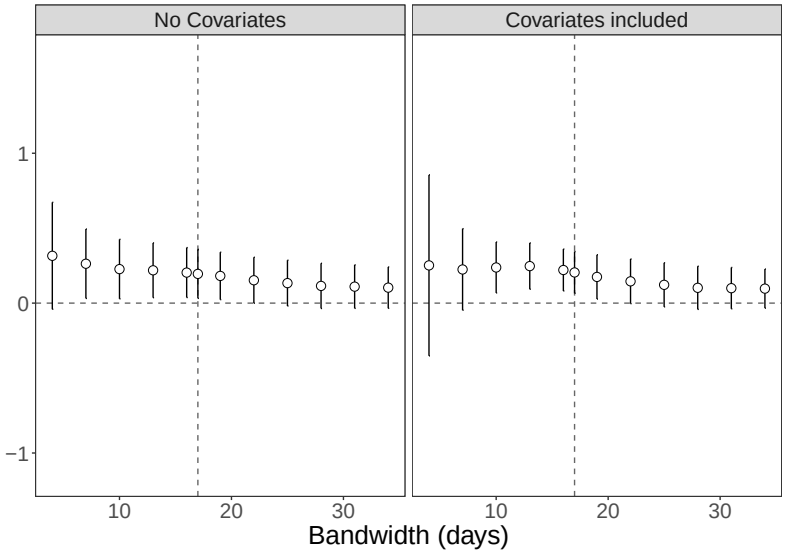
Notes: The Table reports regressions of the cosine similarity between reports and speeches on a continuous time variable capturing the time since exposure. The OLS model includes all controls. In addition, we include fixed effects for the topic of the release. Standard errors for the first model in parentheses are clustered by speech document and by opinion report. In the second model, we include random intercepts for each topic, speech document and opinion report. We also allow for a random slope for each topic. *p<0.1; **p<0.05; ***p<0.01

Figure A.14: Covariate balance around the release of opinion reports



Notes: The figure shows balance of speech-specific covariates around the release of the opinion reports. The results are based on local linear regressions (as specified in Equation 1 in Section 4) around the release of the opinion reports (optimal bandwidth of 22 days).

Figure A.15: RD effects on substantive agreement (bandwidth sensitivity)



Notes: The Figure follows Figure 3 for the substantive agreement outcome.

Table A5: RD effects on substantive agreement for unmatched topics (Placebo)

	Substantive Agreement	
	(1)	(2)
Exposure	0.0261	-0.1053
	(0.0751)	(0.0979)
Covariates	No	Yes
Observations	81	80

Notes: The specification follows Table 6; the sample is limited to pairs where speech document and opinion report do *not* address the same topic.

Table A6: Covariate Documentation

Covariate	**Document**	**Type**	**Description**
Document length	Opinion report	Continuous	Number of words in opinion report document *after* stemming and stopword removal.
Document length	Speech document	Continuous	Number of words in speech document *after* stemming and stopword removal.
Log product of document lengths	Speech document & opinion report	Continuous	Logarithm of the product of document lengths
Year of release	Speech document	Categorical	Year of the release of speech documents
Month of release	Speech document	Categorical	Month of the release of speech documents
Weekday of release	Speech document	Categorical	Weekday of the release
Weekday of release	Opinion report	Categorical	Weekday of opinion report release
Time span between data collection and release	Opinion report	Continuous	Number of days between last day of data collection and date of release of opinion report.
Opinion firm	Opinion report	Categorical	Opinion firm that was responsible for a given report.
Medium of speech	Speech document	Categorical	Medium of speech document as indicated by the German government.

Notes: The Table describes all available covariates used in the empirical analyses. Since the unit of observation is the speech document / opinion report pair, all covariates apply to all units. The column "Document" indicates whether the covariate contains information about a speech document or an opinion report, respectively.