

Hager, Anselm; Hilbig, Hanno

Article — Published Version

Does Public Opinion Affect Political Speech?

American Journal of Political Science

Provided in Cooperation with:

WZB Berlin Social Science Center

Suggested Citation: Hager, Anselm; Hilbig, Hanno (2020) : Does Public Opinion Affect Political Speech?, American Journal of Political Science, ISSN 1540-5907, Wiley, Oxford, Vol. 64, Iss. 4, pp. 921-937,
<https://doi.org/10.1111/ajps.12516>

This Version is available at:

<https://hdl.handle.net/10419/218852>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>

A Online Supplementary Information for: Does Public Opinion Affect Elite Rhetoric?

Anselm Hager and Hanno Hilbig

Contents

A.1	Topic categorization	3
A.2	Classification of untagged speech documents	3
A.3	Topic salience	5
A.4	Substantive agreement	5
A.5	Verbatim quotations	9
A.6	Word contribution	10
A.7	Jaccard Similarity	15
A.8	Parliamentary speech data	16
A.9	Validating Government Public Opinion Research	16

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).

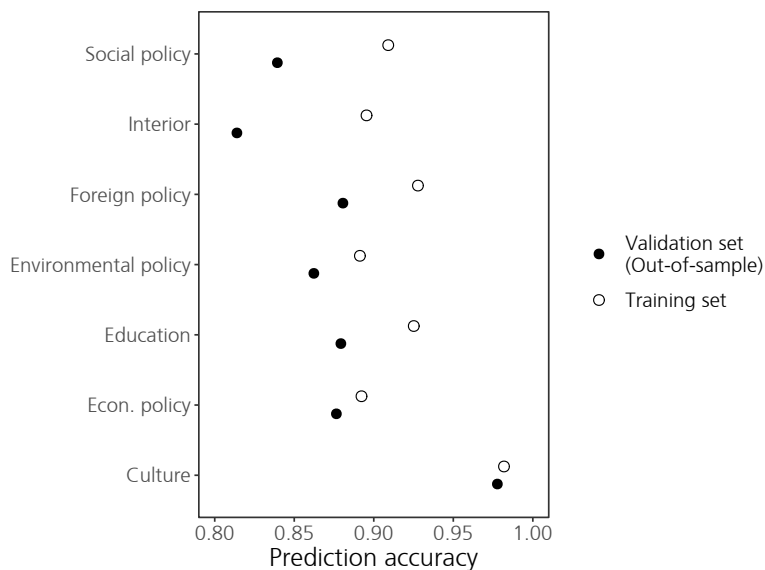
We initially picked two algorithms to predict the topic: The Naive Bayes classifier and the Support Vector Machine (SVM). Here, we focus on the SVM since it performs significantly better than Naive Bayes. To evaluate the performance of the classifier, we choose *prediction accuracy*, i.e., the ratio 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.⁴ 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.

⁴Note 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 $X_{i,k}^*$: $X_{i,k}^* = \frac{X_{i,k}}{\sum_{k=1}^6 X_{i,k}}$.

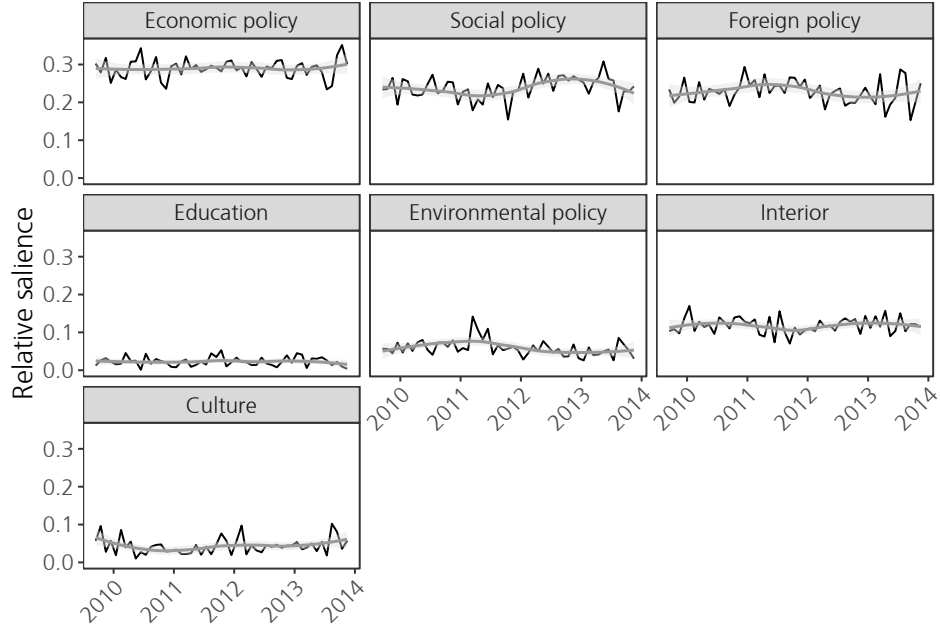
To account for the fact that not all articles are equally relevant, we measured the importance of each article by counting the total number of words in a given article. Let the number of words for article i be l_i . We scaled this measure to range from 0 to 1, such that $l_i^* = F(l_i)$, where F is the 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



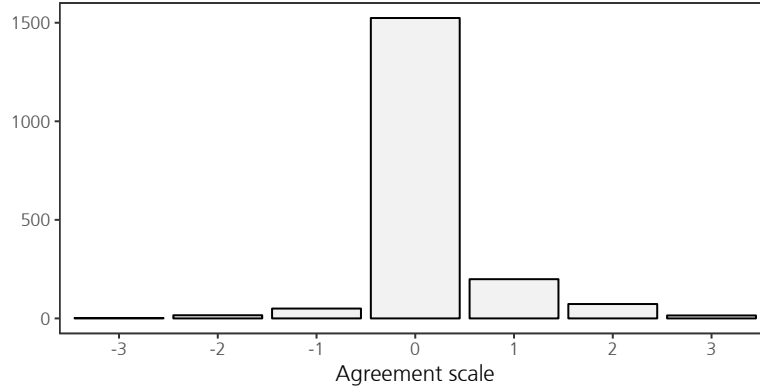
Notes: The Figures plot the relative issue salience per month for the indicated topic (September 2009 to November 2013). The salience measure is constructed from 7,476 articles in *DER SPIEGEL*. Salience denotes the relative number of articles that address an issue. Salience is weighted by the number 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 coders⁵ 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.

⁵The 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



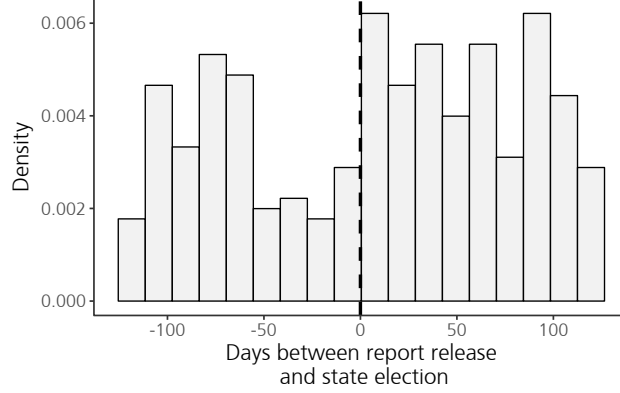
Notes: The plot shows the distribution of the substantive agreement outcome variable, as described in Section A.4 on page 5. Higher values indicate that government speech echoes findings from the opinion reports.

Table A2: Determinants of report timing

Days between data collection and report release							
Salience (SD)	-3.0491 (2.1966)			2.1454 (4.2723)			2.4857 (3.3254)
Report length (SD)		3.6618* (2.0638)			4.1611** (1.9773)		1.7480 (1.3763)
Length of collection			1.5401*** (0.1600)			1.3924*** (0.1657)	1.1389*** (0.1608)
Economic policy				-11.6135** (4.7314)	-7.3402* (4.1203)	-4.6003 (3.2637)	-7.9246** (3.7195)
Foreign policy				-18.0452*** (4.3476)	-15.6711*** (4.7764)	-7.0203* (3.9525)	-8.6014** (3.5531)
Culture				8.6817 (11.9731)	18.8226** (9.1706)	14.6086** (7.2737)	12.2170 (9.4262)
Education				21.0061 (14.3167)	15.1454 (12.7432)	7.7113 (10.1635)	13.6332 (11.0059)
Interior				-3.3927 (8.0477)	-4.7156 (8.4064)	-3.7383 (6.7115)	-0.6613 (6.1902)
Environmental policy				-12.0600 (8.1932)	-10.5010 (7.6611)	-4.3656 (6.1325)	-2.3696 (6.5203)
Constant	24.7805*** (2.1831)	26.0467*** (2.0889)	10.5425*** (2.2577)	36.3204*** (3.5568)	32.1229*** (2.8753)	15.0115*** (3.0438)	18.6569*** (3.6255)
N	82	120	120	82	120	120	82
R ²	0.0235	0.0260	0.4400	0.3131	0.2071	0.4945	0.6109

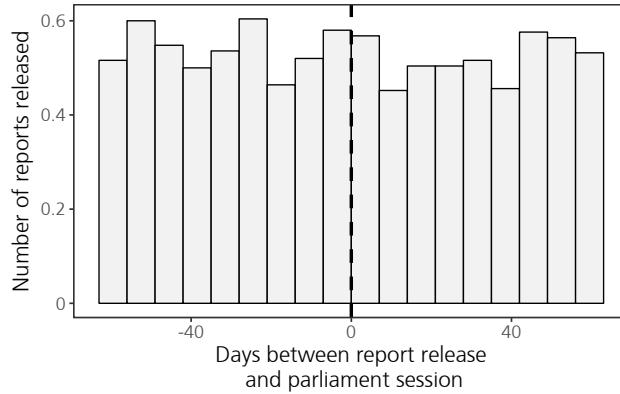
Notes: The dependent variable is the number of days between the last day of data collection and the dissemination date of the report. Topic salience is measured based on a given topic's media salience at the time of the first day of data collection (see Section A.3 on page 5). The length of the report 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



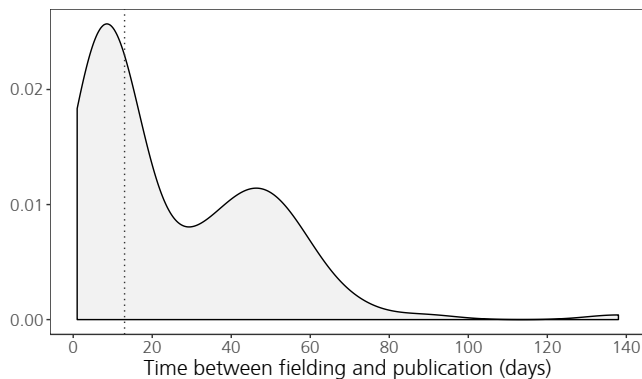
Notes: The Figure plots a histogram of the days between the opinion report release and the date of a state parliament election. Only observations within a window of ± 120 days are included. Negative values on the x-axis indicate that a report was released prior to a state election. The vertical dashed 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



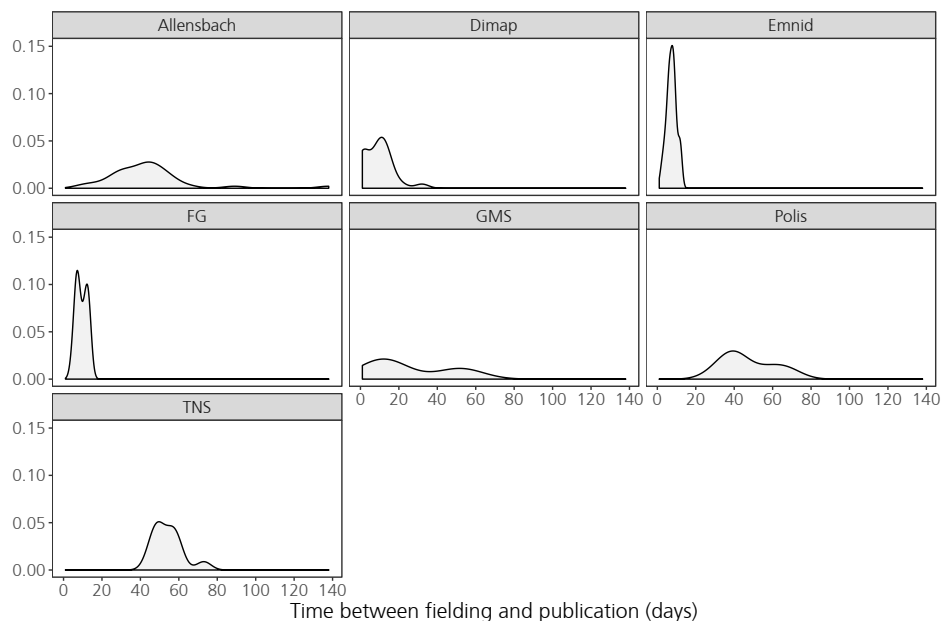
Notes: The Figure plots a histogram of the days between the opinion report release and the days when the parliament is in session. Only observations within a window of ± 60 days are included. Negative values on the x-axis indicate that a report was released prior to a day on which the parliament was 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.

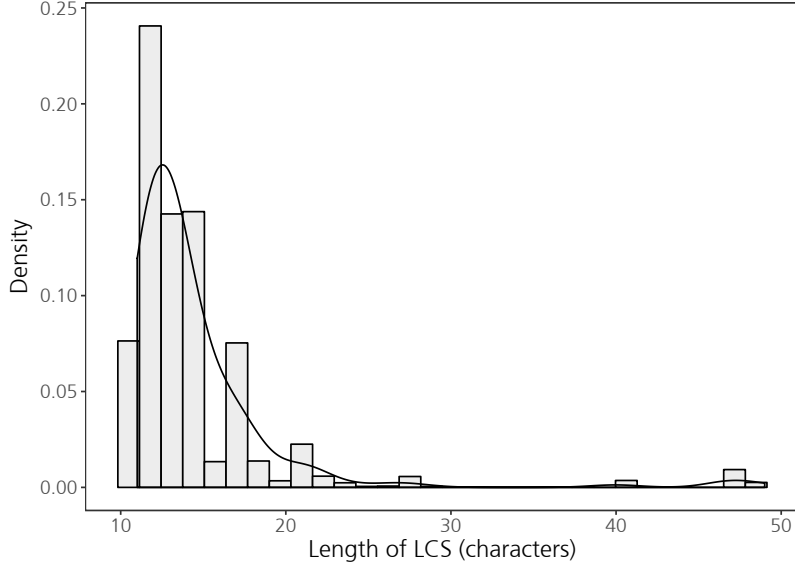
In Figure A.8, we present the distribution of the LCS metric for the set of speech–report pairs where the LCS is longer than 10 characters. As the plot shows, there are very few speech-report pairs for which the longest common substring has more than 20 characters. Given that German words tend to be 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 vectors⁶ 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

⁶The 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.⁷ 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 Δw^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.

⁷We 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]

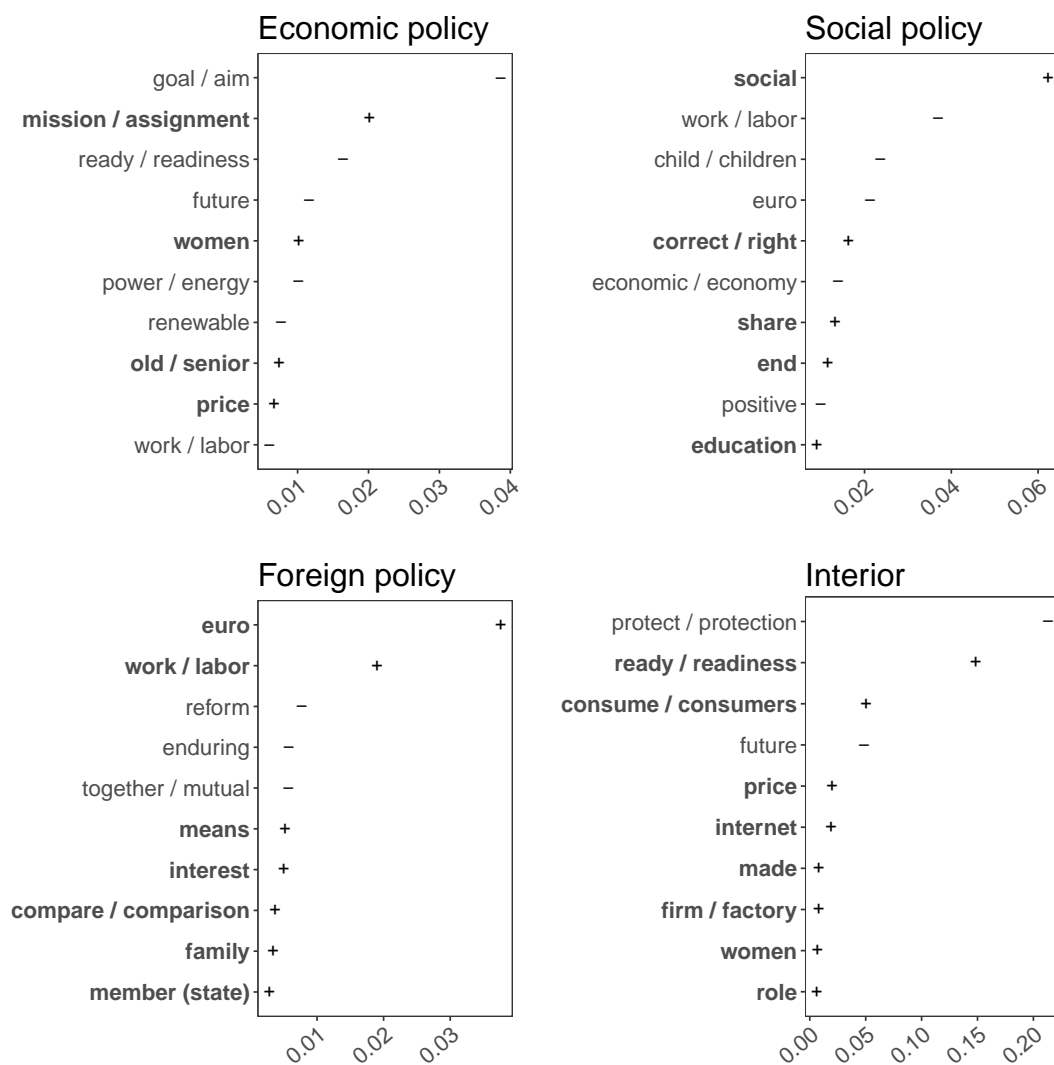
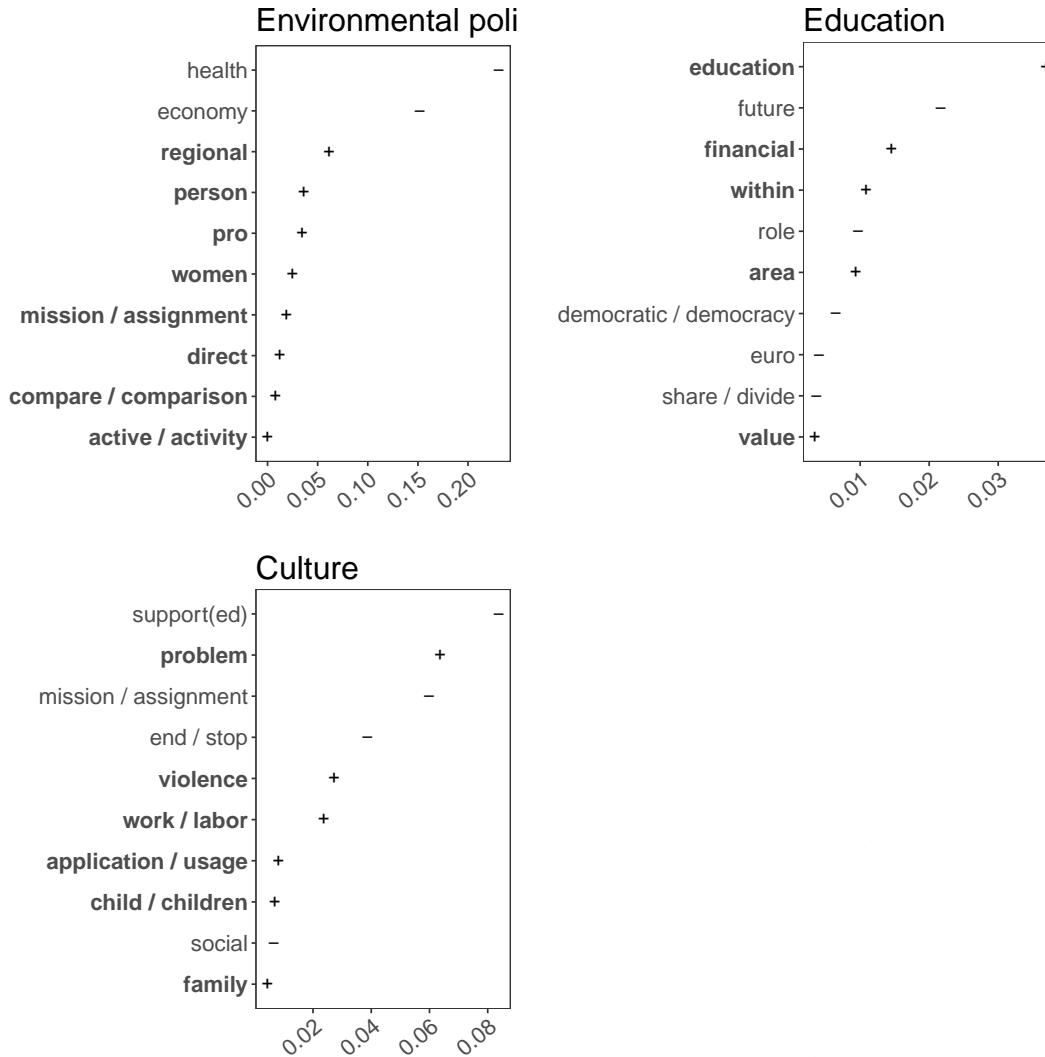


Figure A.12: Word-level contributions to increased cosine similarity [continued]



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

	<i>Jaccard similarity</i>		<i>Individual words</i>	
	(1)	(2)	(3)	(4)
Exposure	0.0153* (0.0084)	0.0102** (0.0036)	3.9612 (2.4921)	1.9998*** (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. $|A_i \cap A_j|$. 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 4.2). 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 3).

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 2.1, 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.

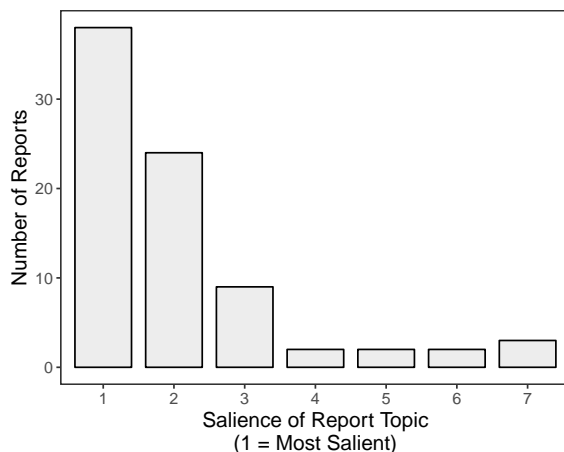
Third, we use the *Politbarometer* surveys to examine whether the government commissions opinion reports as a reaction to their own popularity. The *Politbarometer* surveys ask respondents to state their satisfaction with the work of the federal government on a scale ranging from -5 (very dissatisfied)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

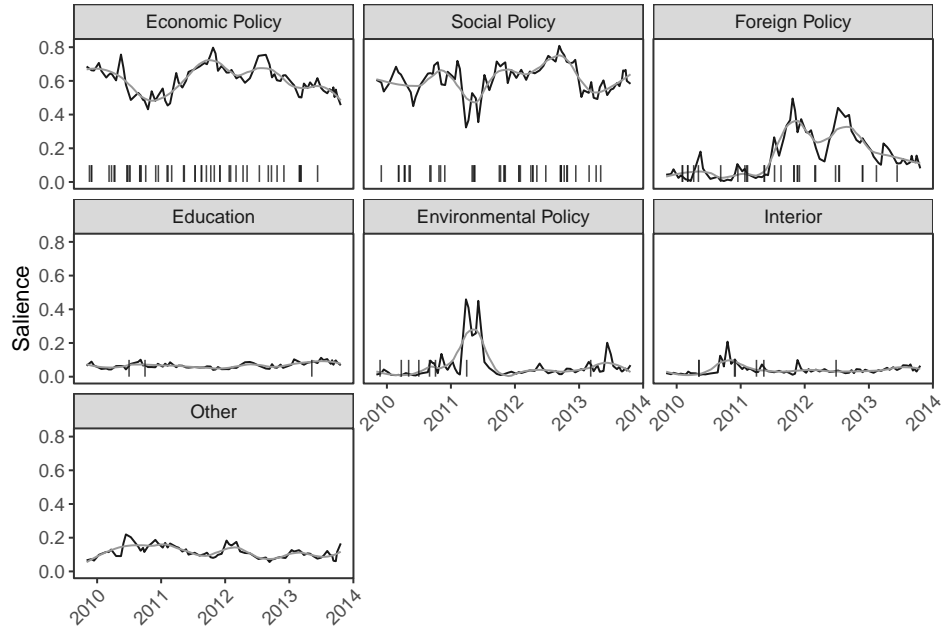
Concluding our analysis, we find that the government broadly commissions reports that cover topics that the public considers to be salient. There is considerable overlap between the issues covered in the Politbarometer and the government opinion research. However, we find little support for the hypothesis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



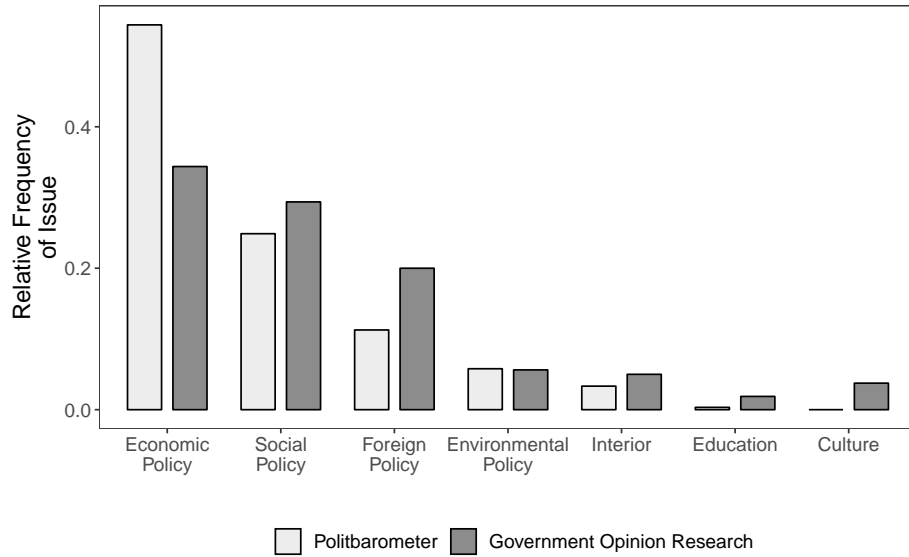
Notes: The Figure shows the number of government reports conditional on the salience of the report topic at the time of data collection ('1' is most salient; salience is based on *Politbarometer* data). The y-axis is the absolute number of reports released.

Figure A.11: Report dates and issue salience



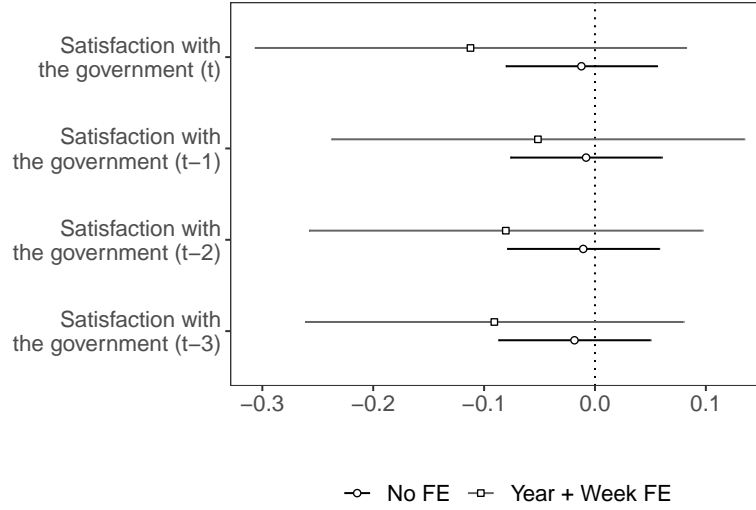
Notes: The panels show issue salience over time based Politbarometer responses. The x-axis is the date, the y-axis indicates issue salience. We plot the raw salience measures (faint grey lines) as well as a LOESS smoother of over-time changes in salience (light and dark blue lines). The vertical lines 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



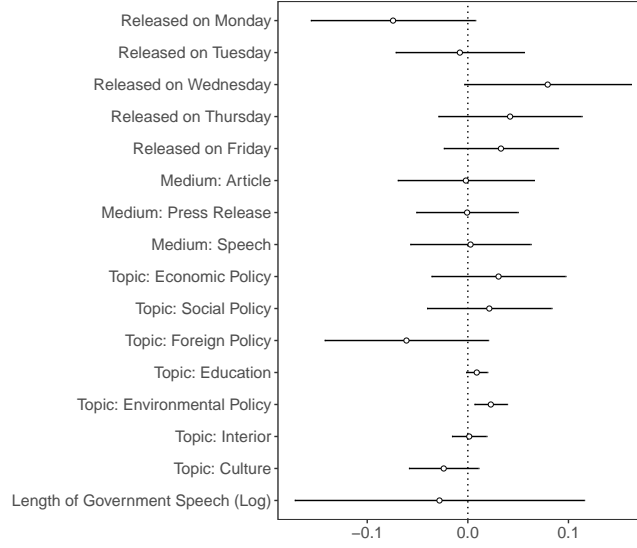
Notes: The Figure shows the effect of a one-standard deviation increase in satisfaction with the government (at different points in time; measured via the *Politbarometer*) on the probability that the government will commission an opinion report. The unit of observation is the week; the outcome is whether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

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

	<i>Cosine similarity</i>	
	OLS	Multilevel Model
Time since exposure (months)	-0.0006** (0.0003)	-0.0008* (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

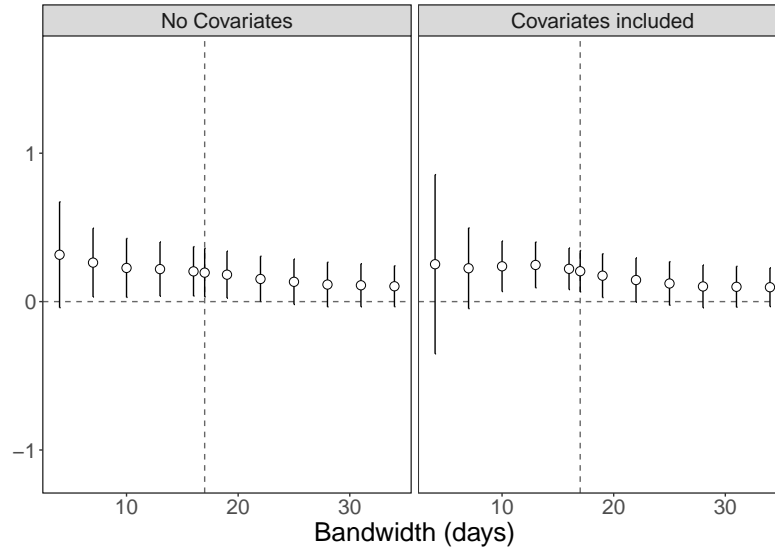
Notes: The Table reports regressions of the cosine similarity between reports and speeches on a continuous time variable capturing the time since exposure. The OLS model includes all controls. In addition, we include fixed effects for the topic of the release. Standard errors for the first model in parentheses 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 3) 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 (0.0751)	-0.1053 (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 <i>after</i> stemming and stopword removal.
Document length	Speech document	Continuous	Number of words in speech document <i>after</i> 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.