Research Article

The International Journal of Press/Politics
2020, Vol. 25(1) 115–134

DOI: 10.1177/1940161219864295
journals.sagepub.com/home/hij
SAGE

Transforming Stability into Change: How the Media Select and Report Opinion Polls

Erik Gahner Larsen[1] and Zoltán Fazekas[2]

Abstract

Although political polls show stability over short periods of time, most media coverage of polls highlights recurrent changes in the political competition. We present evidence for a snowball effect where small and insignificant changes in polls end up in the media coverage as stories about changes. To demonstrate this process, we rely on the full population of political polls in Denmark and a combination of human coding and supervised machine learning of more than four thousand news articles. Through these steps, we show how a horserace coverage of polls about change can rest on a foundation of stability.

Keywords

opinion polls, political coverage, public opinion, media bias

The news coverage in the mass media is decisive for how citizens perceive politics (Ladd and Lenz 2009; Walgrave et al. 2008). An important feature of the political coverage in both nonelection and election times is how parties and politicians stand in the polls. This coverage provides the public and politicians with information on the nature of political competition, shapes the public's attitudes, and matters for political outcomes (Ansolabehere and Iyengar 1994; Rothschild and Malhotra 2014; Searles et al. 2018; van der Meer et al. 2016; Westwood et al. 2018).

The coverage is often dominated by news stories about changes with political reactions and debates about the far-reaching implications of even a single poll. This

[1]Rutherford College, University of Kent, Canterbury, UK
[2]Copenhagen Business School, Frederiksberg, Denmark

Corresponding Author:
Erik Gahner Larsen, Rutherford College, University of Kent, Canterbury, Kent CT2 7NX, UK.
Email: E.G.Larsen@kent.ac.uk

happens despite the fact that most polls show no or little change over short periods of time. As Murray Goot, quoted in Jackman (2005), puts it:

> Fundamentally, the press plays up differences which are otherwise insignificant because it *has* to. Its only alternative is to say that what a poll found today is not significantly different from what it found yesterday; and under most (though not all) circumstances, that sort of news is no news at all. (p. 500, emphasis in original)

Political journalists have to select the most novel political events and decide how to report these events, often times with the imperative to meet a demand for horserace political coverage (Iyengar et al. 2004; Matthews et al. 2012). How can journalists select and report opinion polls to meet the demand for horserace political coverage when such opinion polls rarely provide evidence for novel changes?

Previous studies explaining the coverage of opinion polls fall into two broader categories. First, scholars have studied under what circumstances polls are more likely to be selected by the news media (Groeling 2008; Groeling and Kernell 1998; Searles et al. 2016). Second, a wide range of studies have focused on the characteristics of the reporting of opinion polls (Andersen 2000; Bhatti and Pedersen 2016; Paletz et al. 1980; Weaver and Kim 2002). Both categories of research have provided valuable insights on the coverage of opinion polls, but they relate to theoretically distinct stages of how the media cover opinion polls. We argue that the choices of selecting and reporting opinion polls are related to each other with implications for how individual opinion polls are transformed from descriptive snapshots into newsworthy stories.

To demonstrate this, we study the coverage of opinions polls in a country with a relatively neutral political coverage and no partisan leanings in the media outlets, Denmark (Hallin and Mancini 2004). First, with help from media outlets and polling firms, we tracked down the population of opinion polls from 2011 to 2015 and their media coverage. Second, we used a combination of human coding and supervised machine learning for the content analysis of more than four thousand articles to understand the systematic characteristics of the coverage. We show that polls deviating from previous polls are selected at a higher rate and by more news outlets. This bias toward change travels further into the reporting as well. First, when being reported, polls are highlighted as being about change. Second, even small changes are amplified through political reactions with no rectifications using information about uncertainty.

The Coverage of Opinion Polls

Mass media works as a gatekeeper in relation to which political events should be reported (Groeling and Kernell 1998; Helfer and Aelst 2016; Soroka 2012). Studies have examined how journalists perceive the importance of different events (Stromback et al. 2012), and among the properties that are most relevant to journalists are those revolving around deviations from similar events. The greater the narrative potential for writing about changes related to an event, the more likely it is considered newsworthy (Lamberson and Soroka 2018). As Soroka et al. (2015) write, "[n]ovelty and change are defining features of newsworthiness" (p. 460). In the coverage of the competition between politicians and political parties, this is rooted in the demand for horserace political coverage (Iyengar et al. 2004; Matthews et al. 2012).

Not all political events obtain the same level of attention in the media (Greene and Lühiste 2018; Kostadinova 2017; Meyer et al. 2017). Key to the study of the coverage of political events is defining the population of comparable political events that might be selected. Specifically, we are not able to make inferences related to the coverage of political events vis-à-vis the absence of coverage, often leading to the problem of an unobserved population (Groeling 2013; Hug 2003).

Although previous studies have demonstrated how certain characteristics make events more likely to be selected (Andrews and Caren 2010; Meyer et al. 2017; Niven 2001), we do not fully understand how the focus on change in the horserace coverage can guide both the selection and reporting of political events. When certain events are more likely to be selected, the reporting can accommodate potential biases, for example, by putting emphasis on the unrepresentative nature of the events, but such biases can also be exaggerated and lead to an even more unrepresentative coverage.

Concentrating on vote intention opinion polls provides advantages compared with other political events in the study of horserace political coverage. Opinion polls are seen as newsworthy as they are up-to-the-moment (Paletz et al. 1980) and they come at a high frequency, allowing us to study substantial variation in the coverage of such events. Furthermore, they are quantifiable, thus enabling us to calculate measures of the changes. They are directly comparable as the question wording in the polls on vote intention is identical over time and exogenous to the political context. As opinion polls differ due to random sampling, we are able to study distinct events over time while ruling out confounding factors. Finally, information from opinion polls, such as change and statistical uncertainty, are important features that can be further identified and studied in the reporting.

Some polls are more likely to be selected for coverage by the news media (Groeling 2008; Groeling and Kernell 1998; Matthews et al. 2012; Searles et al. 2016). However, while the literature finds that polls showing greater changes from comparable polls from the same polling firm are expected to be selected, these studies tell us little about how much coverage these polls get and, importantly, how they are being reported on.

In a parallel body of literature, studies have analyzed how journalists present information from polls and concluded that journalists make mistakes in their interpretation of changes between polls (Bhatti and Pedersen 2016; Larson 2003; Toff 2019; Tryggvason and Strömbäck 2018). However, the studies interested in the selection of in-house opinion polls pay limited attention to the reporting and the studies looking at the reporting of political polls do not consider the extent to which the reporting might be related to considerations of the selection of polls.

A Framework to Study Opinion Poll Coverage

Combining the selection and reporting of opinion polls into a single framework can show how both are driven by specific poll characteristics. This allows us to examine the endogenous nature of the opinion poll coverage and provide expectations about what type of coverage we will see.

A poll that is easier to turn into a story about change will be more likely to be selected. The easiest proxy for such potential lies in the property of a poll: does it suggest change and, if so, how large of a change? If this motivation for selecting polls is shared by journalists, it implies an abundant coverage of polls showing large changes. Thus, our first hypothesis is that *polls showing greater change will be selected more often* (Hypothesis 1).

Once an opinion poll is selected for coverage, we expect that there will be a focus on change in the reporting. When looking at the coverage of opinion polls, while the population of opinion polls shows great levels of stability, the media will rather report stories about change in the political competition. Accordingly, we should see that the reporting will be about change despite multiple polls showing stability when taking the margin of error into account (Larson 2003). If a poll was deemed suitable for reporting, the actual magnitude of change is less relevant for the coverage. Independent of the actual change between polls, we will see a focus on change in the reporting. Our second hypothesis is that *change rather than stability reporting is more likely for the selected polls* (Hypothesis 2).

Exploring the Characteristics of Reporting

So far, we argued that there is a discrepancy between the actual change and what the news coverage will suggest in terms of the volatility of party support. It is possible that the reporting will rectify the resulting selection bias, or on the contrary, further amplify it. In this section, we explore ways in which this could be reflected by the content of reporting.

One possible rectification would be dedicating space to the uncertainty these polls come with. As Bhatti and Pedersen (2016) outline, the demand for horserace coverage can lead to unsubstantiated poll stories with errors in the uncertainty reporting (see also Larson 2003). However, we do not know whether the uncertainty reporting is related to the size of the changes between polls.

On one hand, if journalists take the margin of error into account, we could expect that uncertainty will be more likely to be reported for polls showing greater changes. This is because the margin of error will underpin the change story for polls showing significant changes. Furthermore, for smaller changes, the uncertainty might be omitted to not raise doubts about the change narrative. Both these observations would point to a positive relationship between change and uncertainty reporting. However, as we know that polls rarely show significant changes (cf. Bhatti and Pedersen 2016), we can make no strong predictions about such dynamics. On the other hand, if journalists pay greater attention to changes, this can crowd out methodological details such as uncertainty. Accordingly, the margin of error can be one detail of greater relevance for the change narrative when less actual change is present. Thus, this would point toward a negative relationship between change and uncertainty reporting.

Second, journalists can stress the importance of a poll by including responses to the poll from specific sources (Gaskins et al. 2019; Jerit 2009; Tiffen et al. 2014). Specifically, more reactions to a poll indicate higher newsworthiness. Previous research shows that news sources are included in the reporting as means to underscore the political competition and conflicts between parties (Brewer and Sigelman 2002; Dimitrova and Strömbäck 2012). Reactions from experts, commentators, and most importantly politicians confer additional importance to the poll. Such reactions could be more likely to be included when the changes in polls are greater. Moreover, reactions and comments to changes can further feature as independent stories themselves, and such stories are more likely to arise when changes are large and invites reactions. Overall, this would lead us to believe that reporting of polls with larger changes will contain more references and quotes. Alternatively, reporting on polls with smaller changes might require additional validation from external sources to underscore the change narrative. However, this is less consistent with the expectation that any actual change, once a poll is selected, will be reported as change.

In sum, rather than proposing specific hypotheses, we will approach these features of the reporting in relation to change as questions. The answers to these content-related features are important because they can help us better understand whether there is any sort of correction or rather amplification of the reporting. As journalists have different opportunities to use the reporting to rectify or (in)voluntarily amplify biases, we make no strong theoretical predictions about the expected empirical pattern.

Interestingly, some of the mechanisms described above have the potential to add up to a snowball effect. Despite the lack of statistical significance, small changes from one poll to the next are deemed newsworthy and distinct processes can turn polls showing a large degree of stability into a news coverage dominated by stories about recurrent changes in the political competition. In addition, greater changes could affect the level of attention to methodological details as well as political reactions.

Data and Measures

Polls and Change in the Polls

We use the full population of opinion polls in Denmark ($n = 487$) conducted by eight polling firms on vote intention for eight political parties from 2011 to 2015.[1] For information on the parties in the polls in this period, see the Supplementary Information File 1 (SI1). The period covered begins after the 2011 national election and stops prior to the 2015 national election campaign. To ensure that all polls were collected, especially polls not reported, the data set was developed in collaboration with media outlets and polling firms. Denmark, a multiparty Western European democracy, is characterized by a high newspaper circulation and a neutral commercial press (Hallin and Mancini 2004) with extensive reporting of opinion polls (Bhatti and Pedersen 2016).

To measure change in a poll, we rely on the difference between the poll and the last poll from the same polling firm expressed as volatility and measured by the

Pedersen index (Pedersen 1979).[2] The measure—theoretically ranging from 0 (*no change at all*) to 100 (*all previous parties that received support have no support*)—provides a direct measure of change in the political competition in a multiparty system. It is calculated as the sum of gains or losses in absolute terms across all parties and divided by 2.[3]

Identifying Mentions

We collected news articles from nine different newspapers, their webpages and the webpage of two national TV companies.[4] Aggregated, the newspapers had a readership of 1,864,000 (5,643,000 total population) on a normal weekday in the second half of 2014. Four of the newspapers do not have any formal arrangement with a polling firm, whereas the other outlets commissioned polls through the firms used in the analyses.[5]

We collected the news articles using the digital archive from Infomedia, containing all online and print articles in the nation-wide coverage. For each opinion poll, we searched for articles mentioning the polling firm and any party in the articles published within a period of six days after the poll was collected. The time span of six days assures that we focus on the reporting of specific polls as news.

The initial search returned 6,350 articles. We removed all non-Danish reports (a few English language summaries in online editions). Next, we searched the articles for mentions of numbers from the poll to which they were assigned to (such as 16,9, for example, both with decimal comma and point). In addition, we searched for the Danish translation of the bi-grams “new poll” and “new opinion poll.” We define an article as being pertinent if either of these two filters return a positive search result, resulting in a total of 4,147 pertinent articles spread across 412 polls. These steps ensure that we do not include old polls and polls covering other topics than vote intention, such as prime minister preference.

Content Coding

For the content of reporting, we used human coding combined with supervised machine learning. Three research assistants were trained to code five hundred randomly selected articles. Before this coding, we evaluated content coding quality by having a subset of articles coded by multiple coders. Summaries are reported in Table 1 and additional details for the content coding are reported in SI3.

Coders were first shown the title of the article. Titles provide the frame of the article and highlight what has been deemed important. Quantitatively this should be a conservative measurement of reporting change. There might be articles that mention change in the full text, but they do not highlight that in the title, whereas the opposite is unlikely.[6] We code change coverage in the title as 1, if it contains an explicit mention of some actors (parties or party blocks) gaining or losing votes compared with previous opinion polls, 0 otherwise. We found high average inter-coder agreement regarding this feature (88 percent).

Table 1. Human Coding Reliability and Machine Learning Performance.

	Change in Title	Uncertainty	Quote
Human coding			
Coder1:Coder2	40/80/0.52	40/100/1	40/90/0.78
Coder1:Coder3	36/94/0.85	36/94/0.86	36/100/1
Coder2:Coder3	39/92/0.85	39/90/0.69	39/95/0.89
Machine learning			
DFM	174/0.006, 1	422/0.1, 1	1020/0.05, 0.95
Accuracy	0.839	0.977	0.898
Precision	0.875	0.920	0.914
Recall	0.875	0.885	0.939
F score	0.875	0.902	0.926
Data	1643/2254	3343/554	1155/2742

Note. For human coding, entries are number of articles/percent agreement/Fleiss's K. DFM lists number of features (uni- and bi-grams) used, minimum proportion of documents to be present, maximum proportion of documents to be present (sparsity reduction). DFM = document-feature matrix.

The only potential bias favoring confirmation of our hypothesis could appear if the article is not overwhelmingly focused on change while the title mentions change. It is difficult for humans as well as machines to quantify how salient change should be for the article to be only about change. However, most of these articles with change in title actually focus on change and we present empirical checks in SI3 using the full article text where we would predict much higher change reporting. This suggests that the focus on titles do not systematically bias our analysis of the actual reporting.

After coding the title, coders read the article and answered a set of questions related to the content.[7] The two reporting features of interest are (1) whether there was any mention of statistical uncertainty/margin of error of the poll (coded 1 for yes, 0 otherwise), and (2) whether any persons were quoted (coded as 1 for politician [left block or right block], or researcher with university affiliation, or political commentator; 0 otherwise). In both cases, we found high agreement and good reliability (around 95 percent average agreement for both).

Using the labeled text, we apply supervised machine learning to extrapolate human coding to unlabelled articles. The title text corpus was transformed into a document-feature matrix (DFM) used for change in title task, whereas the full content of the articles was used to build the DFM for the uncertainty and quote-related task. We carried out stemming and removed most Danish stopwords with the exception of those that signal directionality (going up or down, for example). We also removed weblinks and punctuation,[8] grouped all numbers into a common token, and carried out the same for party names and party leaders. Although numbers and party names can help with labeling, we limit potential over-fitting to particular names (such as the prime minister) or specific polling numbers. We used both uni- and bi-grams in our analysis.

Using cross-validation, we trained three binary classification models on 80 percent of the labeled documents using gradient boosting (ensemble of decision trees) implemented in `xgboost` (Chen and Guestrin 2016) to best maximize classification accuracy (Olson et al. 2017). All three classifiers yielded good to excellent performance in predicting the binary labels of interest for each document (cf. Table 1). For the title classification, we only used the few words from the title, thus the *F* score is lowest here, but still satisfactory by any conventional standards. We also see good balance between precision and recall, and accuracy levels of 84 percent and above. This means that the machine classifier performance is close to the human coder performance. Specifically, the human inter-coder agreement varies between 88 and 95 percent, while our machine classifier agrees with the humans in around 84 to 87 percent of the cases. Based on this model and the DFM, we predicted labels for the uncoded articles, with the splits in the data reported in Table 1.

Additional Variables

There is some regularity in the timing and frequency of polls, but the intervals between the polls are not constant. These fluctuations can mean longer hiatus of polls or high frequency of polls in a short period of time. To assure that none of these possibilities conflate the results, we control for the distance in days from the last poll by the same polling firm. Although no national election campaign is included in our sample, the period provides variation in the salience of party competition by having campaigns such as the European Parliament elections in 2014 and local elections in 2013. We use a dichotomous variable to control for campaign, taking the value 1 if the poll was released in the three weeks prior to an election, local or European, and 0 otherwise.

Given that our data span across several years, we include a year control in our models. Finally, in subsequent models, we account for potential relationships between polling firms and media outlets using a partnership variable, coded 1 if they had an official agreement.

Results

Polls Showing Change Are Selected More Often

In Figure 1, we summarize changes in polls throughout the period. Although there are two polls with high volatility (discussed in detail in SI1), we see a remarkable stability throughout the period. Beyond stability of the change magnitude, these values indicate small changes. Ninety percent of the polls are below 5.46 percent in terms of volatility, that is, a total of 5 percent electoral support changed in between parties. For reference, between the 2011 and 2015 national elections, this value was 15.97 percent.

Once the margin of error is considered, we can assess whether the changes between two polls for any of the parties were statistically significant.[9] As we are looking at changes between two proximate polls, we find that 82 percent of the polls had no significant changes for any of the parties compared with the previous poll, and 15 percent

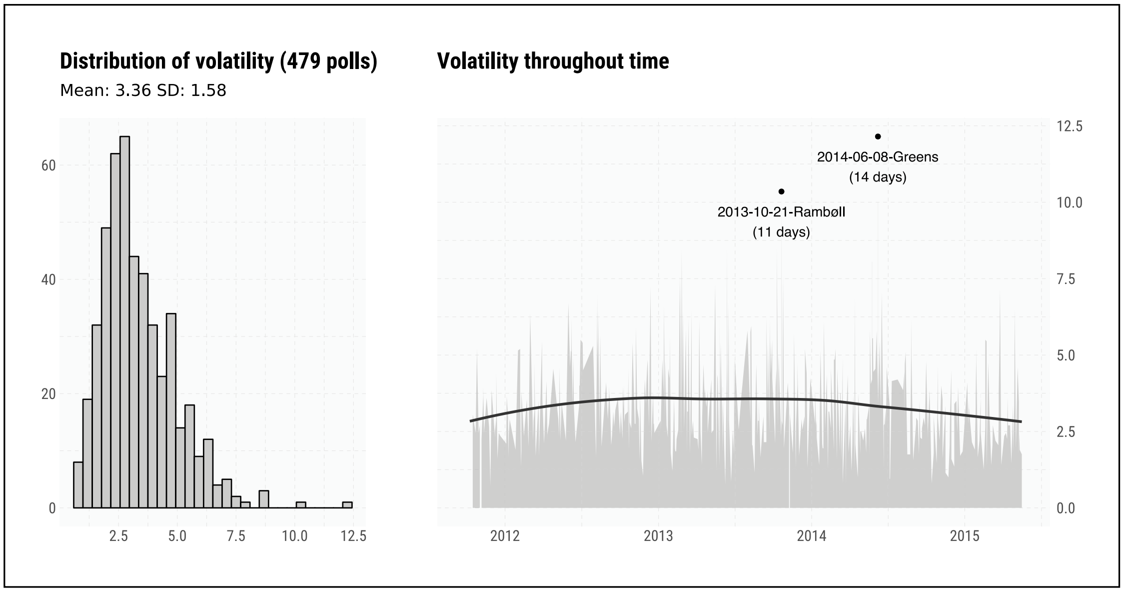


Figure 1. Change in the polls, 2011–2015.
Note. Loess fit overlaid, extreme values highlighted, including distance in days from previous poll released by the same polling firm.

had one or two out of eight potential changes that were statistically significant. Overall, our starting point is a picture of stability in the polls.

In terms of selection, seventy-five polls were not covered, whereas the other polls had mentions ranging from one (forty-one polls) to 131 (one poll) articles.[10] The mention count is summarized in Figure 2. Two polls should be highlighted regarding the high number of mentions. The 2014-05-20-Megafon poll has eighty-eight mentions. This poll falls under the European Union (EU) election campaign period, which we control for. The 2012-05-31-Megafon poll has 131 mentions. This poll attracted additional attention because it listed the Social Democrats—one of the two major parties, also having the prime minister at that point—at a historical low of 16.9 percent.

To test Hypothesis 1, that is, that change translates into higher selection rates, we regressed the count of articles in the media associated with each poll on change and the control variables. Given that the quantity of interest is an overdispersed count, we fitted a negative binomial model to the data. Model 1 in Table 2 summarizes our results and Figure 3 displays the core relationship of interest.

There is systematic variation in the selection of opinion polls. Larger changes are selected more often, indicating that such polls attract more attention from the mass media. We estimate close to twice as many articles (9.41 [7.31, 12.09]) for polls 1 standard deviation above the volatility mean (4.94 percent) compared with those 1 standard deviation below the mean (1.78 percent, 5.72 [4.58, 7.18]).

This ± 1 standard deviation range encompasses 350 polls with the median of the highest gain or loss between polls any party registered being 1.7 percent. Most importantly, in 315 polls (90 percent), there were no statistically significant changes for any of the parties, that is, an overwhelming majority of these changes were within the

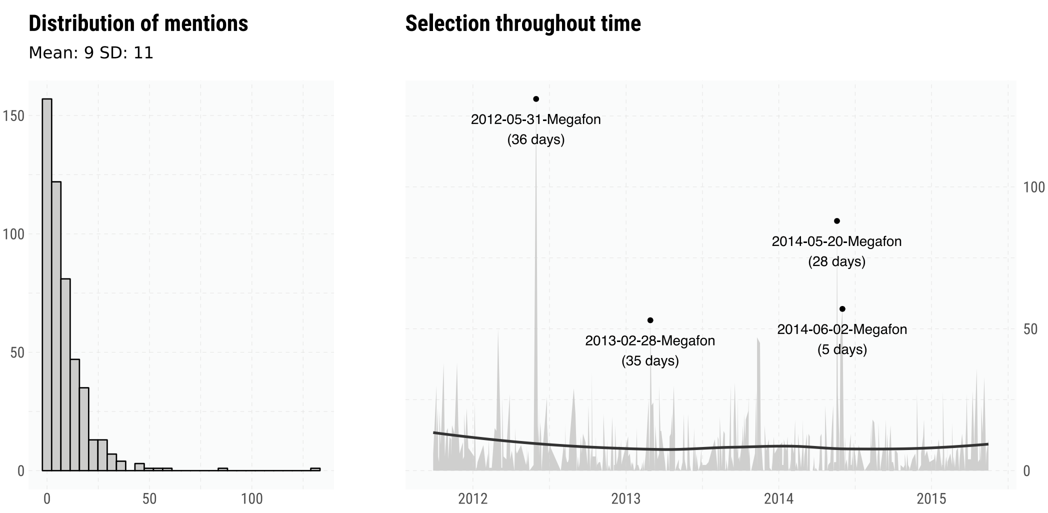


Figure 2. Selection of polls by the media, 2011–2015.
Note. Number of articles (*y*). Loess fit overlaid, extreme values highlighted, including distance in days from previous poll released by the same polling firm.

Table 2. Change and Selection.

	Model 1: Article Count (for Each Poll)	Model 2: Article Count (for Each Poll within Each Outlet)
Change (volatility)	0.50***	0.50***
	(0.14)	(0.05)
Any significant change	0.10	0.14
	(0.17)	(0.07)
Δ days last poll	0.03	−0.11*
	(0.11)	(0.05)
Campaign	1.12***	1.07***
	(0.28)	(0.11)
Partner		1.93***
		(0.31)
Intercept	2.73	−0.33
	(0.24)	(0.14)
Year fixed effects	✓	✓
AIC	3,012	10,622
n	479	5,269
Dyads		88
$\sigma^2_{\text{Intercept}}$		0.89

Note. Bivariate (polyserial) correlation between change and article count is $r = .196$, $p < .001$. Negative binomial regression, restricted maximum likelihood estimates of logit coefficients. For all models, continuous predictors were mean centered and then divided by 2 standard deviations to facilitate comparisons between the magnitude of effects independent of the scale. Standard errors in parentheses. AIC = Akaike information criterion.
*$p < .05$. **$p < .01$. ***$p < .001$.

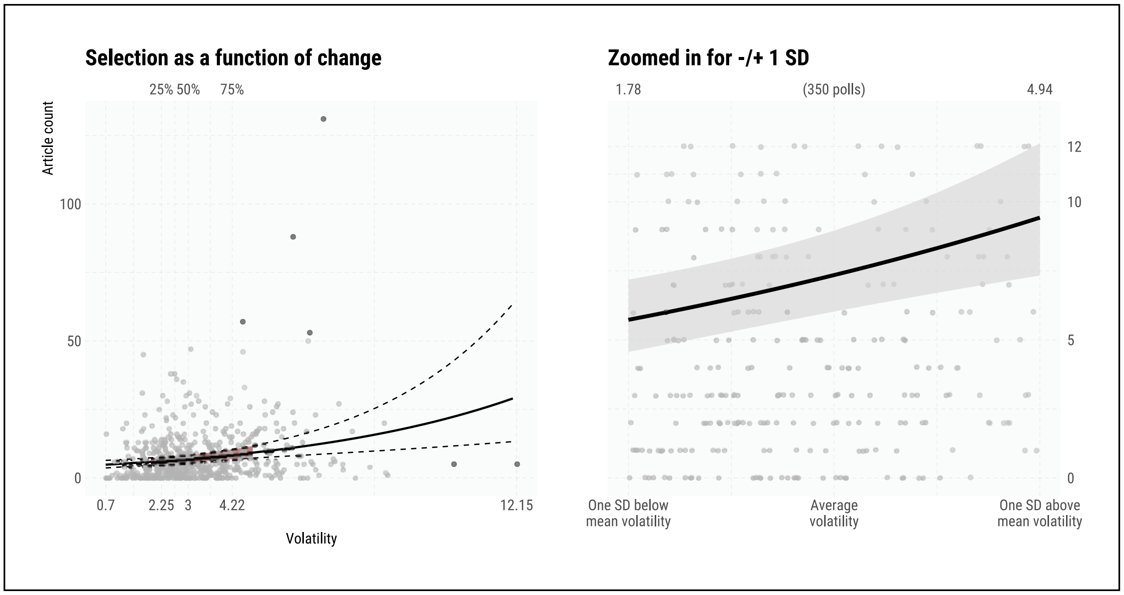


Figure 3. The relationship between change and selection.
Note. Dots represent data with minimal jitter added to aid visualization. Darker dots represent the already introduced cases with high values on the predictor or outcome. In Supplementary Information File 2, we report models with these cases excluded, and our results are identical. Dashed lines (shaded areas in the zoomed plot) are 95 percent confidence intervals. The zoomed plot is a magnified view of a particular range.

margin of error. Finally, this level of volatility (+1 standard deviation above the mean, 4.94 percent) is less than one-third of the actual electoral volatility (15.97 percent) between the 2011 and 2015 general elections, and the difference in volatility (±1 standard deviation, 3.16 percent) is one-fifth of this value. Thus, we see quite some stability, yet, reliance on these differences in limited changes results in twice as many articles.

Next, to account for outlet and polling firm-specific differences, for each poll, we keep the separate counts for the eleven outlets. Thus, we expand our data set, and one poll will have eleven count entries. We account for the hierarchical structure by creating a grouping variable that identifies the combination of the polling firm and the news outlet, a dyad with eighty-eight possible values (8×11). We fit a hierarchical varying intercept negative binomial model to the data and can control now for whether the polling firm and the outlet were partners. The results are reported as Model 2 in Table 2.

There is no evidence for differences across outlets or firms. The positive relationship between change and selection is identical even when we account for outlet- and firm-related heterogeneity. Media outlets select polls that stem from firms with which they have a partnership more often, but we find no evidence that change matter differently in such scenarios.[11]

We make two additional remarks here (described in detail in SI2). First, we find the same relationship between change and selection frequency when refitting the models on the subsample of 392 polls with no significant changes. Accordingly, whether the

Table 3. Change and Reporting.

	Change in Title	Uncertainty	Quote
Change (volatility)	0.10	−1.08***	0.42*
	(0.16)	(0.32)	(0.19)
Any significant change	−0.04	0.29	−0.06
	(0.18)	(0.37)	(0.22)
Δ days last poll	0.01	−1.01***	0.30*
	(0.12)	(0.25)	(0.15)
Campaign	0.07	−0.12	−0.18
	(0.27)	(0.54)	(0.33)
Partner	−0.20*	0.44**	0.35***
	(0.09)	(0.16)	(0.11)
Intercept	0.84	−3.46	1.34
	(0.23)	(0.49)	(0.28)
Year fixed effects	✓	✓	✓
AIC	5,057	2,732	4,441
n	3,822	3,822	3,822
Polls	402	402	402
$\sigma^2_{\text{Intercept}}$	0.58	2.50	0.94

Note. Hierarchical binomial logistic regression, restricted maximum likelihood estimates of logit coefficients. Standard errors in parentheses. AIC = Akaike information criterion.
*$p < .05$. **$p < .01$. ***$p < .001$.

changes are within the margin of error does not matter for the selection practices. Second, for each poll, we looked at the number of different media outlets selecting it and found that, again, change matters in the same way. Polls indicating higher change will spillover to different outlets and not only be disseminated to the readers of the outlet ordering the poll. In sum, we find strong support for Hypothesis 1.

Reporting Change Instead of Stability, Reactions Instead of Uncertainty

In a majority of the cases, the reporting should be about stability. However, 58 percent of the articles mention change in their title. Furthermore, while 82 percent of the polls have no statistically significant changes, 86 percent of the articles does not mention any considerations related to uncertainty.

To test Hypothesis 2 and further examine the content characteristics of the reporting, we employ three hierarchical binomial logit models where *change mention in title, uncertainty mention in text*, and *quotes in text* are modeled as the function of poll characteristics as before.[12] The grouping unit is the poll, as one poll can have multiple associated articles.[13] The results are summarized in Table 3 and Figure 4.

Although there is an increase in the probability of titles highlighting change along the poll change continuum, this effect is not statistically significant and substantively quite small. Thus, the amplification mechanism indicated by our results so far is that

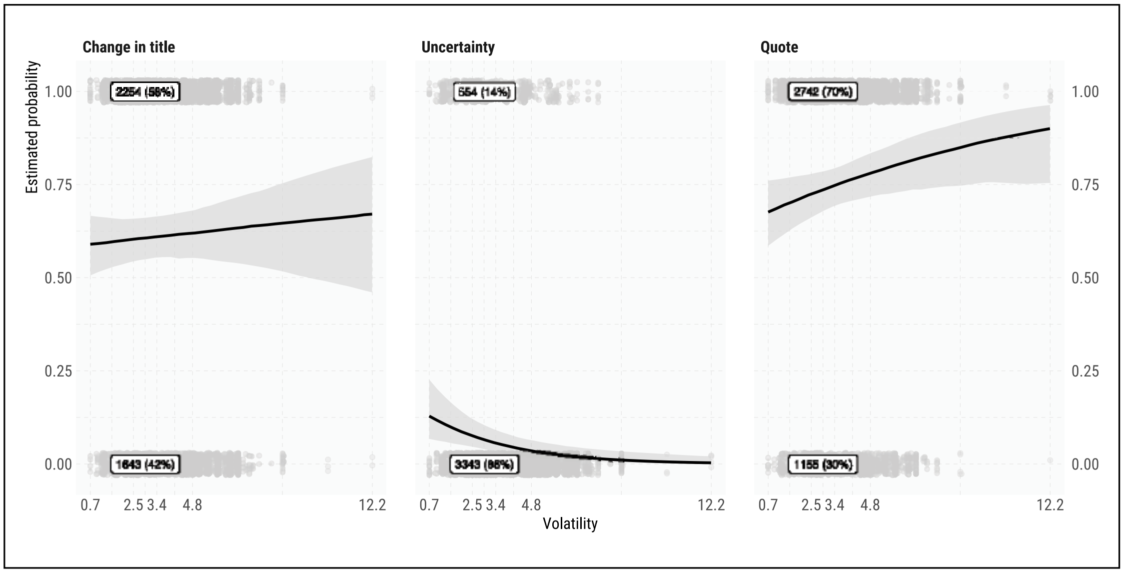


Figure 4. The relationship between change and features of reporting.
Note. Dots represent data with minimal jitter added to aid visualization. Counts and proportions also based on data. Slopes based on hierarchical logit models. Shaded areas are 95 percent confidence intervals.

any volatility level can be sufficient for a change reporting, once selected. In sum, the coverage is dominated by a reporting about change, but this is not exclusive to high change polls, rather a feature of most selected polls. Thus, we find evidence that there is a focus on change rather than stability in line with Hypothesis 2.

Next, interestingly, our results show that reporting of polls with greater change is less likely to make references to uncertainty and the margin of error. Uncertainty is generally not discussed but reporting of high change polls is even less likely to contain such references. Finally, polls showing greater change are reported using more quotes from experts and politicians. Thus, the bias in the selection of the polls travels into fundamental features of the reporting in the form of reactions. In SI3, we report detailed analyses and show that reporting of higher change polls is more likely to quote politicians (effect of 2 standard deviations change is 0.44 [0.19], compared with no quote or other types of quotes). For the content characteristics, the findings demonstrate a pattern where greater changes lead to more political reactions and less focus on the methodological details.

Discussion and Conclusion

The public, politicians, and journalists pay attention to opinion polls as an important source of information (Kerby and Marland 2015; Wichmann and Brettschneider 2009). To understand how such polls are reported in the news media, we have taken an important step forward in understanding the processes polls travel through from their initial collection to their final coverage available to the public. Although polls are often considered newsworthy in nature, we find that there are systematic patterns in how journalists turn these polls into an illusionary political horserace. The overall finding confirms that biases in whether or not polls are covered are not mitigated in the actual reporting but, on the contrary, amplified in the coverage.

We find that selected polls are more likely to be about changes rather than stability. This is despite the fact that most polls show no significant changes. Furthermore, high change poll reporting will focus more on politician voices and commentary and less on important methodological details. As commentators often point out, single outlier polls get a lot of attention. Our results indicate that a thorough consideration and highlight of uncertainty is not part of the coverage that could compensate for such distorted pictures. In other words, the bias in the selection of newsworthy polls is present in the reporting as well.

The findings presented here provide systematic evidence for how the journalistic preference for change has substantial implications for different aspects of the political reporting. Although the model proposed here is useful to make inferences about the political coverage of opinion polls, it can be applied to other types of political coverage. The emphasis on change is not limited to the coverage of polls but matters for the media coverage of other issues such as the economy (Soroka et al. 2015).

There are different ideals for how to report polls as news. One possibility is to report everything, independent of changes and including all information, or potentially reporting only those polls where the indicated changes are significant. Although there are other ideal expectations, these two example scenarios would satisfy the idea of unbiased reporting, but these would generate very different representations of political competition. The media cannot devote attention to all polls, but even if they could, the public demand for horserace coverage would lead to some polls getting more attention than others, especially in competitive media systems.

The dynamics illustrated above have substantial implications for contemporary democracies. Opinions polls have been shown to be the single most important predictor of people's expectations of parties' electoral success (Zerback et al. 2015) and can affect the perceived closeness of parties (Cushion et al. 2016). Such perceptions have implications for not only whether people will show up and vote on election day but also for what types of parties that people vote for, especially when voting strategically. When news articles focus on notional changes in public opinion, such changes can have substantial political implications. Although we do not examine the effects of such an emphasis on changes, we see important questions for future inquiry related to the behavioral and attitudinal implications of these dynamics.

The findings presented here corroborate findings in studies from other countries, including the United States. However, some caveats are important to keep in mind. First, the data are collected in a context with limited polarization and largely nonpoliticized news media. We cannot rule out that the effects can be further amplified when additional motivations are included in the framework. For example, the coverage of opinion polls in Denmark from 2011 to 2015 was chosen to provide a relatively homogeneous set of media outlets with no partisan connections or leanings. Although the journalistic preference for change is present in all media systems, future research will have to examine the conditional nature and relative relevance of the preference for change vis-à-vis other motivations. In addition, while the Danish setting outside a context of an election campaign provides a conservative test of our argument as the horserace coverage is less prominent here, additional evidence is needed to substantiate the generalizability of the findings beyond the period studied here.

Finally, our data does not allow us to analyze the mechanisms of why journalists interview sources and avoid reporting information on the margin of error. For the margin of error, Pétry and Bastien (2013) suggest that journalists follow the interpretations put forward by the polling firms. However, if journalists mainly rely on materials from their partner organizations, we should not see the widespread and outlet independent finding at both stages.

Although the increase of public opinion polls has decreased the uncertainty politicians deal with when assessing the state of public opinion (Geer and Goorha 2003), somewhat ironically this decreased uncertainty comes with a potentially stronger bias in the coverage. As shown in the paper, this intertwined process highlights an ever-growing challenge for political journalists. They need to meet the demand for horserace journalism while selecting and reporting opinion polls in a factual and representative manner.

Authors' Note

Both authors contributed equally.

Acknowledgments

We thank Maja Laundrup Christensen, Maria Clara Verdich, and Joachim Jan Pulawski Legh-Smith for excellent research assistance. For valuable feedback, we thank Lars Gylling, Morten Skovsgaard, Martin Ejnar Hansen, Federico Vegetti, Sebastian Popa, Rasmus Tue Pedersen, Yosef Bhatti, Erik Albæk, Kim Andersen, Bjørn Høyland, Arjen van Dalen, the editor, and three anonymous reviewers at the *International Journal of Press/Politics*.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The text coding was funded by the Departmental Grant Scheme at the Department of Political Science and Public Management, University of Southern Denmark.

ORCID iD

Erik Gahner Larsen https://orcid.org/0000-0003-3579-8457

Supplemental Material

Supplemental material for this article is available online.

Notes

1. These are Epinion, Gallup, Greens, Megafon, Rambøll, Voxmeter, Wilke, and YouGov. Norstat carried out five similar polls in this period. To ensure that we use polls that are comparable, that is, conducted on a regular basis, we do not include these five polls.
2. Although for five polling firms there is information available on polls prior to the starting date of our period covered, these are all pre-2011 national election numbers; hence, we do not include them here, resulting in the drop of the first poll in our period of data coverage in terms of volatility calculation. Using either election results or those previous polls does not change our results.
3. In Supplementary Information File 4 (SI4), we report all our analysis using a different change measurement, that is, the greatest magnitude of change between two polls. Our findings are unchanged.
4. These are *Berlingske, Børsen, B.T., Jyllands-Posten, MetroXpress, Politiken, Ekstra-Bladet, Kristeligt Dagblad*, and *Information*, with webpages b.dk/politiko.dk, borsen.dk, bt.dk, jyllands-posten.dk, mx.dk, politiken.dk, ekstrabladet.dk, kristeligt-dagblad.dk, and information.dk. The TV companies are the Danish Broadcasting Company (dr.dk) and TV 2 (politik.tv2.dk).
5. One polling firm has a news outlet partner not included directly in our data collection, as they are commissioned by Ritzau, the largest Danish independent news agency. One outlet (Jyllands-Posten) switched firms in 2013, from Rambøll to Wilke. We account for this partnership agreement in our models, see below. Throughout the analysis, we keep online and offline platforms from the same media outlet as one outlet.
6. To keep a close match with what our coders coded we will focus on the title text only. However, we also carried out checks where for supervised learning we used the title and the first three sentences of an article (potential headline summary) and the results reported below are unchanged.
7. As we randomized from the original 6,350 documents, some were nonpertinent, hence, lower sample size than fifty for cross-coded articles.
8. All text-related work was carried out using `quanteda` (Benoit 2018) and `tidytext` (Silge and Robinson 2016) in R. We picked sparsity reductions empirically, running the classifier on a range of varying lower and upper inclusion criteria.
9. Following Bhatti and Pedersen (2016), we approach this problem as difference of proportions between two independent polls and use a .05 significance level. In our models, we will employ a dichotomous variable that takes the value 1 if any of the changes were statistically significant and 0 otherwise.
10. Detailed descriptive statistics are presented in SI1.
11. We tested whether the effect of volatility varies across dyads and that is not the case. We report a more detailed discussion in SI2.
12. We have consciously avoided testing our argument using a two-stage selection model. Our main claim is that change, a property of the poll, together with other poll features will influence both stages. Accordingly, the only possible identification strategy we could employ is a functional one. We also have a discrepancy between the potential number of observations, with one poll having multiple articles at the outcome stage, making such a model even more difficult to identify. Most importantly, our theoretical argument implies that outcome-related considerations (content coverage) factor into the selection stage, rather than selection-related features exerting influence on the outcome level.

13. This also implies that polls reported only once will have no within-group variation. For such cases, our outcome estimates will be pulled toward the grand mean (i.e., probability of title mentioning change).

References

Andersen, Robert. 2000. "Reporting Public Opinion Polls: The Media and the 1997 Canadian Election." *International Journal of Public Opinion Research* 12 (3): 285–98.

Andrews, Kenneth T., and Neal Caren. 2010. "Making the News: Movement Organizations, Media Attention, and the Public Agenda." *American Sociological Review* 75 (6): 841–66.

Ansolabehere, Stephen, and Shanto Iyengar. 1994. "Of Horseshoes and Horse Races: Experimental Studies of the Impact of Poll Results on Electoral Behavior." *Political Communication* 11 (4): 413–30.

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

Bhatti, Yosef, and Rasmus Tue Pedersen. 2016. "News Reporting of Opinion Polls: Journalism and Statistical Noise." *International Journal of Public Opinion Research* 28 (1): 129–41.

Brewer, Paul R., and Lee Sigelman. 2002. "Political Scientists as Color Commentators: Framing and Expert Commentary in Media Campaign Coverage." *International Journal of Press/Politics* 7 (1): 23–35.

Chen, Tianqi, and Carlos Guestrin. 2016. "XGBoost: A Scalable Tree Boosting System." In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://www.kdd.org/kdd2016/papers/files/rfp0697-chenAemb.pdf.

Cushion, Stephen, Richard Thomas, Allaina Kilby, Marina Morani, and Richard Sambrook. 2016. "Interpreting the Media Logic behind Editorial Decisions: Television News Coverage of the 2015 U.K. General Election Campaign." *International Journal of Press/Politics* 21 (4): 472–89.

Dimitrova, Daniela V., and Jesper Strömbäck. 2012. "Election News in Sweden and the United States: A Comparative Study of Sources and Media frames." *Journalism* 13 (5): 604–19.

Gaskins, Ben, Jason Barabas, and Jennifer Jerift. 2019. "Qualitative Quotes: The Prevalence and Effects of Survey Respondent Exemplars in Political News Coverage." *International Journal of Press/Politics*. Published electronically June 3. doi:10.1177/1940161219851071.

Geer, John, and Prateek Goorha. 2003. "Declining Uncertainty: Presidents, Public Opinion and Polls." In *Uncertainty in American Politics*, ed. Barry C. Burden, 139–60. Cambridge: Cambridge University Press.

Greene, Zachary, and Maarja Lühiste. 2018. "Symbols of Priority? How the Media Selectively Report on Parties' Election Campaigns." *European Journal of Political Research* 57 (3): 717–39.

Groeling, Tim. 2008. "Who's the Fairest of Them All? An Empirical Test for Partisan Bias on ABC, CBS, NBC, and Fox News." *Presidential Studies Quarterly* 38 (4): 631–57.

Groeling, Tim. 2013. "Media Bias by the Numbers: Challenges and Opportunities in the Empirical Study of Partisan News." *Annual Review of Political Science* 16 (1): 129–51.

Groeling, Tim, and Samuel Kernell. 1998. "Is Network News Coverage of the President Biased?" *Journal of Politics* 60 (4): 1063–87.

Hallin, Daniel C., and Paolo Mancini. 2004. *Comparing Media Systems: Three Models of Media and Politics*. Cambridge: Cambridge University Press.

Helfer, Luzia, and Peter Van Aelst. 2016. "What Makes Party Messages Fit for Reporting? An Experimental Study of Journalistic News Selection." *Political Communication* 33 (1): 59–77.

Hug, Simon. 2003. "Selection Bias in Comparative Research: The Case of Incomplete Data Sets." *Political Analysis* 11 (3): 255–74.

Iyengar, Shanto, Helmut Norpoth, and Kyu S. Hahn. 2004. "Consumer Demand for Election News: The Horserace Sells." *Journal of Politics* 66 (1): 157–75.

Jackman, Simon. 2005. "Pooling the Polls over an Election Campaign." *Australian Journal of Political Science* 40 (4): 499–517.

Jerit, Jennifer. 2009. "Understanding the Knowledge Gap: The Role of Experts and Journalists." *Journal of Politics* 71 (2): 442–56.

Kerby, Matthew, and Alex Marland. 2015. "Media Management in a Small Polity: Political Elites' Synchronized Calls to Regional Talk Radio and Attempted Manipulation of Public Opinion Polls." *Political Communication* 32 (3): 356–76.

Kostadinova, Petia. 2017. "Party Pledges in the News: Which Election Promises Do the Media Report?" *Party Politics* 23 (6): 636–45.

Ladd, Jonathan McDonald, and Gabriel S. Lenz. 2009. "Exploiting a Rare Communication Shift to Document the Persuasive Power of the News Media." *American Journal of Political Science* 53 (2): 394–410.

Lamberson, P. J., and Stuart Soroka. 2018. "A Model of Attentiveness to Outlying News." *Journal of Communication* 68 (5): 942–64.

Larson, Stephanie Greco. 2003. "Misunderstanding Margin of Error Network News Coverage of Polls during the 2000 General Election." *The International Journal of Press/Politics* 8 (1): 66–80.

Matthews, J. Scott, Mark Pickup, and Fred Cutler. 2012. "The Mediated Horserace: Campaign Polls and Poll Reporting." *Canadian Journal of Political Science* 45 (2): 261–87.

Meyer, Thomas M., Martin Haselmayer, and Markus Wagner. 2017. "Who Gets into the Papers? Party Campaign Messages and the Media." *British Journal of Political Science*. Published electronically September 29. doi:10.1017/S0007123417000400.

Niven, David. 2001. "Bias in the News: Partisanship and Negativity in Media Coverage of Presidents George Bush and Bill Clinton." *International Journal of Press/Politics* 6 (3): 31–46.

Olson, Randal S., William La Cava, Zairah Mustahsan, Akshay Varik, and Jason H. Moore. 2017. "Data-Driven Advice for Applying Machine Learning to Bioinformatics Problems." *arXiv:1708.05070*.

Paletz, David L., Jonathan Y. Short, Helen Baker, Barbara Cookman Campbell, Richard J. Cooper, and Rochelle M. Oeslander. 1980. "Polls in the Media: Content, Credibility, and Consequences." *Public Opinion Quarterly* 44 (4): 495–513.

Pedersen, Mogens N. 1979. "The Dynamics of European Party Systems: Changing Patterns of Electoral Volatility." *European Journal of Political Research* 7 (1): 1–26.

Pétry, François, and Frédérick Bastien. 2013. "Follow the Pollsters: Inaccuracies in Media Coverage of the Horse-Race during the 2008 Canadian Election." *Canadian Journal of Political Science* 46 (1): 1–26.

Rothschild, D., and Malhotra, N. 2014. Are public opinion polls self-fulfilling prophecies? *Research & Politics*. https://doi.org/10.1177/2053168014547667

Searles, Kathleen, Martha Humphries Ginn, and Jonathan Nickens. 2016. "For Whom the Poll Airs: Comparing Poll Results to Television Poll Coverage." *Public Opinion Quarterly* 80 (4): 943–63.
Searles, Kathleen, Glen Smith, and Mingxiao Sui. 2018. "Partisan Media, Electoral Predictions, and Wishful Thinking." *Public Opinion Quarterly* 82 (Suppl. 1): 302–24.
Silge, Julia, and David Robinson. 2016. "tidytext: Text Mining and Analysis Using Tidy Data Principles in R." *The Journal of Open Source Software* 1 (3): 37.
Soroka, Stuart N. 2012. "The Gatekeeping Function: Distributions of Information in Media and the Real World." *Journal of Politics* 74 (2): 514–28.
Soroka, Stuart N., Dominik A. Stecula, and Christopher Wlezien. 2015. "It's (Change in) the (Future) Economy, Stupid: Economic Indicators, the Media, and Public Opinion." *American Journal of Political Science* 59 (2): 457–74.
Stromback, Jesper, Michael Karlsson, and David Nicolas Hopmann. 2012. "Determinants of News Content: Comparing Journalists' Perceptions of the Normative and Actual Impact of Different Event Properties When Deciding What's News." *Journalism Studies* 13 (5–6): 718–28.
Tiffen, Rodney, Paul K. Jones, David Rowe, Toril Aalberg, Sharon Coen, James Curran, Kaori Hayashi, Shanto Iyengar, Gianpietro Mazzoleni, Stylianos Papathanassopoulos, Hernando Rojas, and Stuart Soroka. 2014. "Sources in the News: A Comparative Study." *Journalism Studies* 15 (4): 374–91.
Toff, Benjamin. 2019. "The 'Nate Silver Effect' on Political Journalism: Gatecrashers, Gatekeepers, and Changing Newsroom Practices around Coverage of Public Opinion Polls." *Journalism* 20 (7): 873–89.
Tryggvason, Per Oleskog, and Jesper Strömbäck. 2018. "Fact or Fiction? Investigating the Quality of Opinion Poll Coverage and Its Antecedents." *Journalism Studies* 19 (14): 2148–67.
van der Meer, Tom W. G., Armen Hakhverdian, and Loes Aaldering. 2016. "Off the Fence, onto the Bandwagon? A Large-Scale Survey Experiment on Effect of Real-Life Poll Outcomes on Subsequent Vote Intentions." *International Journal of Public Opinion Research* 28 (1): 46–72.
Walgrave, Stefaan, Stuart Soroka, and Michiel Nuytemans. 2008. "The Mass Media's Political Agenda-Setting Power: A Longitudinal Analysis of Media, Parliament, and Government in Belgium (1993 to 2000)." *Comparative Political Studies* 41 (6): 814–36.
Weaver, David, and Sung Tae Kim. 2002. "Quality in Public Opinion Poll Reports: Issue Salience, Knowledge, and Conformity to AAPOR/WAPOR Standards." *International Journal of Public Opinion Research* 14 (2): 202–12.
Westwood, Sean J., Solomon Messing, and Yptach Lelkes. 2018. "Projecting Confidence: How the Probabilistic Horse Race Confuses and Demobilizes the Public." Working paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3117054.
Wichmann, Wolfgang, and Frank Brettschneider. 2009. "American and German Elite Journalists' Attitudes toward Election Polls." *International Journal of Public Opinion Research* 21 (4): 506–24.
Zerback, Thomas, Carsten Reinemann, and Angela Nienierza. 2015. "Who's Hot and Who's Not? Factors Influencing Public Perceptions of Current Party Popularity and Electoral Expectations." *International Journal of Press/Politics* 20 (4): 458–77.

Author Biographies

Erik Gahner Larsen is a lecturer in Quantitative Politics at the School of Politics and International Relations at the University of Kent. His current research focuses on the interconnections between public policies, public opinion, and the media. His previous work has been published in outlets such as the *European Sociological Review, British Journal of Political Science*, and *Journal of Elections, Public Opinion & Parties*.

Zoltán Fazekas is an associate professor of Business and Politics, with focus on quantitative methods at the Copenhagen Business School, Department of International Economics, Government and Business. He studies political behavior at the voter and the elite level. His research is at the intersection of political psychology, political communication, and comparative politics. His previous work has been published in outlets such as the *American Journal of Political Science, Journal of Communication*, and *International Journal of Public Opinion Research*, among others.

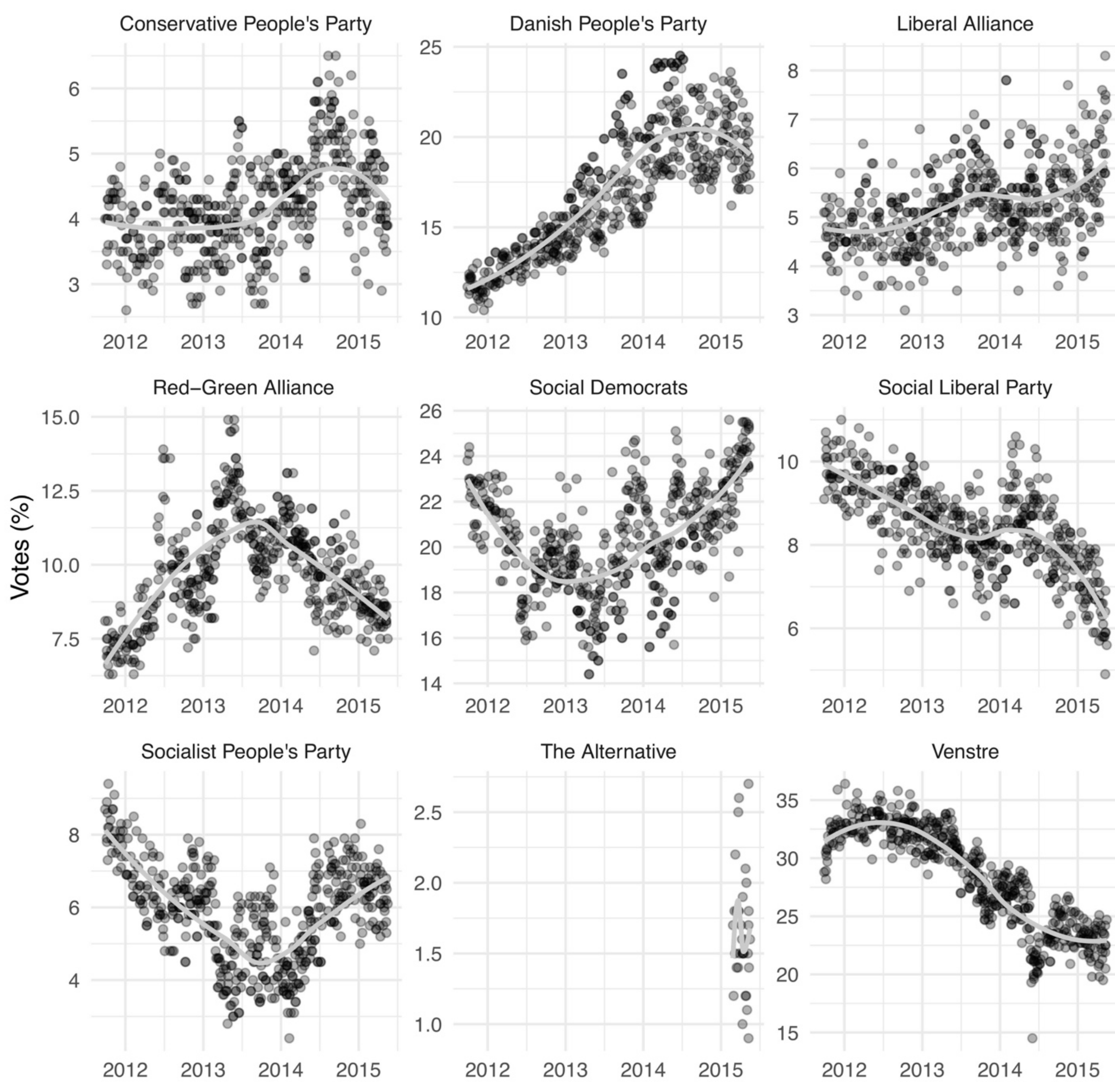
Supplementary Information:

"Transforming Stability into Change: How the Media Select and Report Opinion Polls".

1 Context and descriptive statistics

In the aftermath of the Danish general election in 2011, the Social Democrats, the Social Liberal Party and the Socialist People's Party formed a three-party coalition government with the support of the Red-Green Alliance. Figure A1.1 shows the support for nine political parties measured throughout this period.

Figure A1.1: Support for the parties, 2011-2015



In the period, there was a new entrant in Danish politics, The Alternative. The polling firms disagreed on the level of support for this party with some polls having them above the electoral threshold of 2% and other below, although a majority of these poll numbers were not statistically significant from each other. Given that this party has much fewer measurements as a new party, we do not include them in the main analysis.

The three government parties lost support during the first two years in office, partially due to a series of unpopular reforms and broken pledges. Some of the voters switched to parties within the red block, i.e. went to the Red-Green Alliance, whereas others went to the blue parties such as the Danish People's Party and Venstre. Towards the end of this period, the Social Democrats regained support in the public and Venstre lost a substantial number of votes and polled below 25% in a majority of the polls towards the 2015 general election. However, most importantly, none of these changes happened from one single poll to the next, and the year controls included in the reported models take any potential effects of these dynamics into account.

Table A1.1 shows descriptive statistics on the polls for each polling firm, including the number of polls, the average number of mentions in the coverage and the average volatility.

Table A1.1: Detailed descriptive statistics

Firm	Total polls	Earliest	Latest	Average mentions	Average volatility	Average significant changes	Days btw. polls
Epinion	39	2011-11-03	2015-05-12	6	3.105	0.18	34
Gallup	45	2011-10-06	2015-05-07	10	3.511	0.43	30
Greens	42	2011-10-06	2015-04-29	6	4.543	0.66	32
Megafon	49	2011-10-06	2015-04-30	22	4.224	0.35	27
Rambøll	75	2011-10-13	2014-09-15	4	4.242	0.59	14
Voxmeter	144	2011-09-28	2015-05-17	9	2.353	0.03	9
Wilke	18	2013-11-10	2015-05-10	7	3.732	0.29	32
YouGov	75	2011-10-10	2015-05-11	5	3.193	0.03	18

As noted in the main text, two polls (Rambøll, 2013-10-21 and Greens, 2014-06-08) were outliers in terms of volatility. In the Rambøll poll, the Social Democrats lost 5.4 percentage points and the Danish People's Party gained 4.7 percentage points. This was not a trend followed by other polling firms and it was an outlier poll for this firm as well. In the Greens poll, the Social Democrats gained 8.2 percentage points and Venstre lost 7 percentage points. Here, no other polling firms showed similar trends.

Last, in order to look into the context in which polls were selected and covered in more detail, we first examined the five most and five least volatile polls and whether they were selected or not. Second, we looked into the coverage of polls quoting sources and mentioning uncertainty for polls with low and high volatility.

For the polls covered, the highest volatility is 12.15% (as in total change in vote share estimate, Greens 2014-06-08), whereas for those not covered this value is 7.6% (Greens, 2013-10-25). Overall, we found no outliers in extreme polls not being covered, as all polls with the most extreme volatility were indeed covered. The most extreme polls covered and not-covered come from two polling firms, Rambøll and Greens. In other words, the most volatile non-covered and covered polls are from the same firms, confirming that our main selection results are not explained by systematic differences in whether a poll is picked up conditional on the polling firm.

For the polls with the smallest volatility covered and not covered, we see much less variation. This is explained by the fact that many polls show little variation. For lowest volatility polls we see an over-representation of polls from Voxmeter. While they are a large part of our population, the consistent weekly frequency of polls from them leads to only minor differences for a lot of the polls, some of which gets covered. One of the polls getting coverage is a poll with the story that a party, Venstre, gets "exactly 33.3 percent of the votes" (Voxmeter, 2011-11-27). Interestingly, the polls with the lowest volatility getting covered are framed as no change (e.g. Voxmeter, 2014-11-16 and 2014-11-23) or in relation to specific events that makes the support for the parties relevant, e.g. op-eds (Voxmeter, 2013-08-18) or the European Parliament election (Epinion, 2014-04-21).

For the polls with high volatility, we see that experts comment on the changes. The poll with the most coverage, Megafon (2012-05-31), resulted in articles where expert comments can be summarized as the changes were unique and that it had implications for the next parliamentary elections, with politicians commenting on the changes in the same way. A similar example is the poll from Rambøll (2013-10-10), that also resulted in pundits and politicians reacting to the poll due to the changes. Accordingly, we see that for the polls where there is a high degree of volatility, sources are included to comment on the changes, and in particular the causes of these changes and their potential consequences.

For the polls with the low volatility, once they get covered and there are reactions from sources, the comments vary from being about the support for a *specific* political party or potential future implications of these changes. In other words, there are cases when sources are included in low volatility polls, but they are less frequent and with no systematic pattern in what they comment on, in contrast to the overwhelming change focus in high volatility poll coverage.

2 Models of selection: additional details

Here we centralized all additional models, alternative specifications, and detailed considerations related to our argument about how change is associated with more frequent selection.

2.1 Alternative measures and specifications

In Tables A2.1 and A2.2 we report three models in each table, the only difference being that in the first table we have the robustness checks for the specification where we look at each poll as unit of analysis and have one overall media article count, whereas the second table reports models in which each poll has eleven (outlet) entries in terms of counts, with outlet × polling firm grouping and a hierarchical setup.

Table A2.1: Alternative models: negative binomial models of overall article count

	(1) Extreme values excluded	**(2)** Not significant only	**(3)** Maximum change
Intercept	2.72***	2.79***	2.71***
	(0.23)	(0.24)	(0.24)
Change (2 SD)	0.45**	0.57***	0.49***
	(0.14)	(0.17)	(0.14)
Any significant change (= 1)	−0.07		0.04
	(0.17)		(0.18)
Days since last poll (2 SD)	0.03	0.02	0.08
	(0.11)	(0.12)	(0.11)
Election campaign (= 1)	0.98***	1.29***	1.09***
	(0.28)	(0.30)	(0.28)
2012	−0.68**	−0.73**	−0.59*
	(0.25)	(0.26)	(0.26)
2013	−0.71**	−0.72**	−0.71**
	(0.25)	(0.26)	(0.26)
2014	−0.94***	−1.04***	−0.92***
	(0.25)	(0.26)	(0.26)
2015	−0.54*	−0.64*	−0.55
	(0.28)	(0.29)	(0.28)
AIC	2931.59	2401.98	3014.04
LogLik	−1455.80	−1191.99	−1497.02
N	473	392	479

Notes: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Year = 2011, baseline.

Model (1) in both scenarios lists results when we exclude the extremely high volatility polls (above 10%) identified and discussed in the paper, but also four polls that overall had more than 50 mentions. In the hierarchical setup the per outlet article number is lower, hence there we only exclude the high volatility polls. Model (2) in both cases keeps only those polls that had no statistically significant changes compared to their previous counterparts, whereas finally, Model (3) substitutes volatility as a measure of change with the maximum change registered by a party for each poll.

Table A2.2: Alternative models: hierarchical negative binomial models of article count

	(1) Extreme values excluded	(2) Not significant only	(3) Maximum change
Intercept	−0.15	−0.11	−0.20
	(0.15)	(0.16)	(0.15)
Change (2 SD)	0.53***	0.51***	0.46***
	(0.06)	(0.06)	(0.06)
Any significant change (= 1)	0.14		0.12
	(0.07)		(0.08)
Days since last poll (2 SD)	−0.12**	−0.05	−0.09
	(0.05)	(0.05)	(0.05)
Election campaign (= 1)	1.06***	1.25***	1.03***
	(0.11)	(0.11)	(0.11)
2012	−0.68***	−0.85***	−0.60***
	(0.11)	(0.11)	(0.11)
2013	−0.71***	−0.77***	−0.66***
	(0.11)	(0.11)	(0.11)
2014	−1.09***	−1.20***	−1.06***
	(0.11)	(0.11)	(0.11)
2015	−0.65***	−0.72***	−0.60***
	(0.12)	(0.12)	(0.12)
AIC	10622.59	8390.63	10677.16
LogLik	−5300.30	−4185.31	−5327.58
N	5247	4312	5269
Dyads (outlet × company)	88	88	88
Var (Intercept)	1.22	1.26	1.22

Notes: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Year = 2011, baseline.

From an empirical point of view, as highlighted in the main text, these results bring further evidence that our main findings are not contingent on couple of extreme observations (1) or on the choice of change measurement (3). From a substantive point of view, we also showed that change matters even when the change is “illusionary” (2), as in statistical uncertainty related to polling estimates would suggest stability.

2.2 Media outlet count

To further our understanding of amplification beyond the number of articles, we also regressed the number of different outlets out of the 11 total outlets that report on a particular poll on the predictors employed in the paper. In Table A2.3 we report the results from two models: in the first column all polls are included (including those with no reporting, i.e. 0 outlet count), whereas in the second column we subset our data to only those polls that received at least one mention. In both cases, we find evidence for amplification of reporting of change through diversification of the outlets reporting: polls indicating more change will be picked up by more different outlets.

Table A2.3: Model results: negative binomial model of outlet count

	Outlet count	Outlet count (one or more reports)
Intercept	1.84^{***}	1.80^{***}
	(0.17)	(0.12)
Change (2 SD)	0.38^{***}	0.19^{*}
	(0.10)	(0.08)
Any significant change (= 1)	−0.09	−0.01
	(0.13)	(0.10)
Days since last poll (2 SD)	−0.00	−0.07
	(0.08)	(0.06)
Election campaign (=1)	0.47^{*}	0.35^{*}
	(0.20)	(0.15)
2012	-0.58^{**}	-0.41^{**}
	(0.18)	(0.13)
2013	-0.56^{**}	-0.32^{*}
	(0.18)	(0.13)
2014	-0.76^{***}	-0.48^{***}
	(0.18)	(0.13)
2015	-0.46^{*}	-0.35^{*}
	(0.20)	(0.15)
AIC	2248.77	1874.26
LogLik	−1114.39	−927.13
N	479	404

Notes: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Year = 2011, baseline.

2.3 Hierarchical model extension

As referenced in the main text, we also fitted a model where the effect of volatility is let to vary across dyads, with correlation across varying effects fixed estimated. Model fit comparison

(Table A2.4) indicates that the effect of change does not vary across the outlet and polling firm combinations.

An alternative way to specify this model would be to use polls as the grouping level (or, level-2) and have a uniform 11 observations within each group, with volatility and other poll related quantities treated as level-2 predictors. However, this would mean that for all polls that were not reported, we would have no within-group variation, i.e. all outcome values would be 0. While this is not a fundamental issue, it would also mean that the only level-1 predictor is whether the outlet and the polling firm were partners. Letting this effect vary across polls and potentially adding a model to this variation with volatility being a predictor (cross-level interaction), would enable us to discuss whether the positive effect of partnership varies as a function of volatility. We can reach similar substantive conclusions regarding the effect of partnership as in our main approach, however, that relationship is not the core quantity or predictor of interest here.

Table A2.4: Model fit comparison: varying effect of change (across outlet × firm dyads)

	Df	AIC	BIC	deviance	Chisq	Chi Df	p-value
Original model	12	10623	10701	10598.62			
Varying slope of change	14	10626	10718	10597.93	0.69	2	0.7090

3 Reporting: additional details

3.1 Title coding task

The title coding was carried out in two steps, i.e. two questions:

(1) does it contain any mention of a party or party-block support (yes or no), if *yes*:
(2) what kind of support interpretation is given, with the options: (1) close race, (2) status quo, standstill, (3) party/block names is losing votes, (4) party/block names is winning votes, (5) one party/block is winning, another losing.

In the main text, we treat answers 2.3, 2.4, and 2.5 as (1 – there is change mention in the title), all the rest of the options, including no support mention (no for question 1) as 0. This is to provide a conservative test where we might underestimate the actual focus on change in the reporting. We have re-run our analysis using an alternative coding where close race is also coded as change. Inter-coder reliability is unchanged, the supervised machine learning results presented in the main text are slightly better as "close race" tends to mention parties or blocks as does change coverage, but all substantive results are the same. In other words, decisions related to how close race should be treated are not influential for our results.

3.2 Classifier summary

Once training the classifiers, for each particular categorization task, we can summarize which features (uni- or bi-grams build from stems) carry importance in assuring accurate classification. The terms listed in Table A3.1.1 and Table A3.1.2 (translated) are the top 40 most important terms for each of the three main classifiers (usually 110-130 features with non-0 importance). Again, these are not "directional" *per se* (although can be added based on what class of documents they appear more often). Instead, they indicate that if the texts contain these features, the classifier will do better in differentiating between the classes. While we do not list the exact gains in terms of prediction error reduction associated each feature, it is worth noting that the top 5-10 range carries the meaningful weight.

Table A3.1.1: Importance ranked terms

	Change in title	Uncertainty	Quote		Change in title	Uncertainty	Quote
1	måling	usikkerh	siger	21	i_ny	måske	ifølg
2	meningsmål	tokenanynumn_procentpoint	sagd	22	valget	fire	op_til
3	blok	procent	ritzau	23	røde	inden	målinger
4	sender	statistisk	politisk_ordfører	24	tokenanynumn	ritzau	kan
5	mellem	tokenanynumn	så	25	chokmål	blandt_andet	procent
6	nedtur	er_tokenanynumn	politisk	26	radikal	ved	udvalgt
7	ny	derfor	komment	27	tager	tidliger	tilbagegang
8	katastrofemål	repræsentativt	partiet	28	frem	gør	største
9	går	står	tror	29	historisk	politik	i_stedet
10	ved	foretaget	lyder	30	dag	opbakn_fra	lige
11	vælgern_stemm	den_tokenanynumn	valgforsk	31	store	over	bruge
12	politisk	person	gik	32	stormer	tilbagegang	til_tokenanynumn
13	får	vore	ordfører	33	mandat	i_folketinget	blandt
14	større	nogensind	universitet	34	fremgang	procent_af	politik
15	flertal	i_tokenanynumn	fordel	35	siden	senest	stor
16	spærregrænsen	meget	rød	36	lige	niveau	statsministeren
17	tilbag	godt	og_så	37	største	stadig	på_tokenanynumn
18	analys	langt	uger	38	vælgere	kommer	ting
19	giver	vælgern	gå	39	vælgern	svarer	står
20	regeringen	partiet	dansk	40	fast	landet	danskern

Table A3.1.2: Importance ranked terms (translated)

	Change in title	Uncertainty	Quote		Change in title	Uncertainty	Quote
1	poll	uncertain	says	21	in_new	maybe	according
2	opinion poll	tokenanynumn_percentagepoints	said	22	election	four	up_to
3	bloc	percentage	ritzau	23	red	before	polls
4	sender	statistical	political spokesman	24	tokenanynumn	ritzau	can
5	between	tokenanynumn	saw	25	shock	among_other	percentage
6	downturn	is_tokenanynumn	political	26	radical	by	selected
7	new	therefore	comment	27	takes	former	decline
8	disaster poll	representative	party	28	forward	do	largest
9	going	stands	believes	29	historical	political	instead
10	by	conducted	says	30	Day	support_from	equal
11	voter_vote	the_tokenanynumn	researcher	31	big	over	use
12	political	person	went	32	storms	decline	to_tokenanynumn
13	gets	our	spokesman	33	mandate	in_parliament	among
14	larger	ever	university	34	progress	percentage_of	political
15	majority	in_tokenanynumn	advantage	35	since	latest	large
16	threshold	very	red	36	equal	level	primeminister
17	back	good	and then	37	largest	still	on_tokenanynumn
18	analys	Long	weeks	38	voters	coming	thing
19	provides	voters	walk	39	voter	similar	stands
20	government	party	danish	40	firm	landed	danes

3.3 From title change to change in text

As highlighted in the main text, there are operational and theoretical reasons to focus on the title text rather than the full text of the article when assessing the change mentions in the reporting. Two additional empirical considerations underline that there is no systematic bias in favour of confirming our second hypothesis. In the results in the main text we find no evidence for a relationship between change in titles and change in polls, but we do find that change in title is a frequent part of the reporting, i.e. even small volatility polls are presented as change once they get through the selection change.

In our first approach, we randomly selected 40 articles, equally split between labelled as title having change or no change. An additional coder blind to the selection and goal coded whether change was mentioned in the content of each of the articles, based on reading the full text. Next, in a separate file (differently ordered), the coder was asked to do the same based on the title text alone, mimicking the task in our main sections of the paper. According to this coding step, the proportion of change mentions coded based on titles was 0.4 (compared to 0.5 in the data), but the proportion was 0.775 based on the coding of the full text. This indicates that through the coding of titles alone, we are likely underestimating the change reporting compared to what we were to get based on coding the full text. When we subset our extra coding dataset, we see that these numbers are 0.75 for articles where original title coding was done by other coders and 0.8 for those where the labelling is the result of the machine learning.

To reiterate, the main aim here is not to explicitly think of correspondence between coding based on titles *vs* full text, rather to see in which direction the differences appear. In this regard, while a limited exercise, 31 out of the 40 articles were coded to have change mentions *based on the full text,* and in our data 17 of these are labelled as not being about change *based on the titles*. Overall, we saw good human coding and machine learning performance based on titles and those are used for our main analysis; if we were to think of reporting features based on full article texts, we should expect that change coverage is even more often mentioned.

In our second approach, we modify the prediction steps of our machine learning approach to further substantiate that full text based change coding could only strengthen our claim that change is ubiquitous in the media reporting about opinion polls. As a first step, similar to our main analysis, we trained our classifier using the title texts. We trained two classifiers, one with no reduction in terms of number for features (9,990 features in total) and one that yielded the best performance (174 features).

For the random 20% subset of our data (*test set, not included in training)*, we used the trained classifier to predict labels of change or no change. To do this, we used the document-feature matrix created according to the same rules of the full texts. To be more precise, we used the features that were present in the training set (titles) and also in the full texts. However, these features came with different frequencies. When we used the full text on the *test set* prediction, the average proportion of change was 0.877 (using all features, no sparsity reduction) and 0.872 (reduced sparsity, features present in at least 24 titles) respectively. On the same test sets, if we use the title based document-feature matrix and the same classifier, the proportion change was 0.64 and 0.63. Thus, the first take-away, consistent with the human coding exercise, is that full texts would only indicate higher change reporting. If we conduct any transformation (such as *tf-idf*) to account for the length of full texts, these numbers will be even higher, above 0.9. In terms of interpretation, this simply indicates that full texts contain words associated with change derived from the titles, and proportionally they carry an important weight. It does not necessarily mean that 80% of the articles are only about change, but they do make enough change references (more than one) to be regarded as change reporting.

As human coders worked with title texts for this task and the coding is based on that, accuracy on the test set using title texts is much higher than that using full text (0.839 *vs* 0.645). Furthermore, likely both change and stability vocabulary is larger for full texts. What is noteworthy is that out of 75 no change labels by the human coders in the test set, the prediction based on titles mislabels 17 as being about change, but this is 65 for the case full text content. This final piece of information is to underscore that using titles to capture change reporting is unlikely to bias the findings upwards, i.e. to overestimate the amount of change reporting.

3.4 Detailed consideration of quote sources

To provide additional information on the analysis of quoted sources, we keep the original coding of quotes with minor reduction of complexity only. Specifically, we do not differentiate between red and blue block politicians (treated as Politician) and between university affiliated experts or political commentators (treated as Expert). We follow the same procedure as before for human coding and machine learning.

The inter-coder reliability and agreement for the 4-category[1] quote measure was in line with other numbers reported in the main text, i.e. very good: 90%/0.86 (coder1:coder2), 89%/0.84 (coder 1:coder 3), 92%/0.88 (coder 2:coder 3). For the supervised machine learning the 0.1 lower threshold was deemed best, resulting in a dfm with 422 features. A multinomial classifier with boosting was used, resulting in good performance given the difficulty of the task: 0.82 accuracy and confusion matrix reported in Table A3.2. F1 scores for each category in order were: 0.84, 0.86, 0.79, and 0.70. After labelling the full dataset, 32% had no quote, 37% quoted a politician, 19% an expert, and 12% of articles quoted both.

Table A3.2: Confusion matrix

	Observed			
	0	1	2	3
Predicted 0	52	7	3	0
1	7	70	1	1
2	2	3	34	1
3	2	3	9	20

We fitted four hierarchical models (with grouping at the poll level) for the 4-category quote split up. We do not fit a hierarchical multinomial model because its complexity and the cases of no-within poll variation (only one report per for a poll) create estimation difficulty. The results are reported in Table A3.3.

Two consistent findings emerge: as volatility increases, there is a sharp drop of "No quote" scenario compared to all other options together, and an increase in the probability of quoting a politician compared to all other options together. We also find that change is not associated with

[1] 0 = No quote, 1 = Politician quoted, 2 = Expert quoted, and 4 = Both quoted.

expert quoting. Finally, while quotes covering multiple types of sources are more likely with higher change polls, these differences are also not statistically significant.

Table A3.3: Hierarchical models of quote type

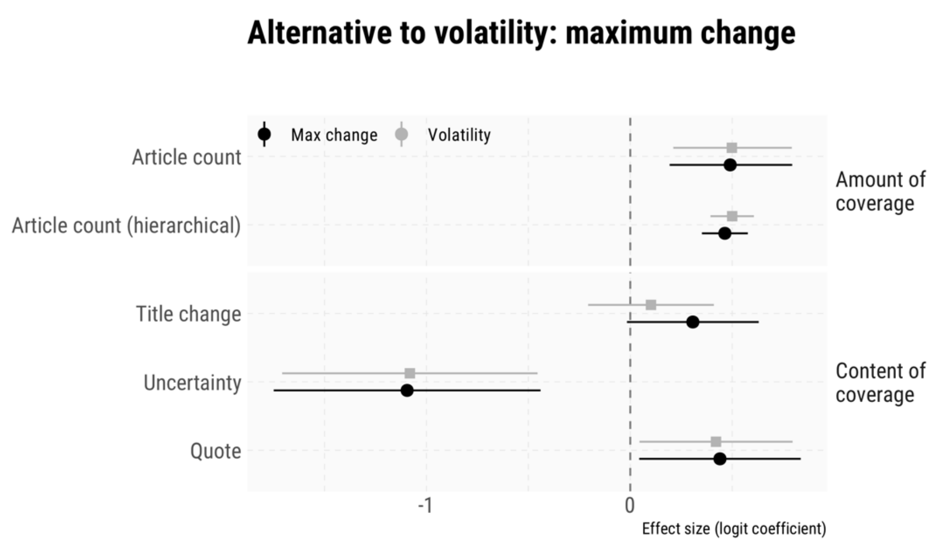
	No quote	Politician quote	Expert quote	Both quoted
Intercept	-1.47***	-0.23	-2.42***	-1.99***
	(0.28)	(0.27)	(0.35)	(0.38)
Change (2 SD)	-0.51**	0.44*	0.03	0.28
	(0.19)	(0.19)	(0.22)	(0.28)
Any significant change (= 1)	0.03	0.15	-0.41	0.29
	(0.22)	(0.22)	(0.26)	(0.31)
Days since last poll (2 SD)	-0.27	0.09	-0.08	0.55**
	(0.14)	(0.14)	(0.17)	(0.21)
Election campaign (= 1)	0.08	-0.03	0.28	0.13
	(0.32)	(0.34)	(0.37)	(0.49)
Partner (= 1)	-0.31**	-0.02	-0.10	0.68***
	(0.10)	(0.10)	(0.12)	(0.13)
2012	0.60*	-0.73*	1.04**	-0.68
	(0.30)	(0.30)	(0.37)	(0.42)
2013	0.43	-0.45	0.81*	-0.61
	(0.30)	(0.30)	(0.37)	(0.42)
2014	0.80**	-0.67*	0.86*	-1.09*
	(0.30)	(0.30)	(0.37)	(0.43)
2015	0.99**	-0.73*	1.15**	-1.94***
	(0.33)	(0.34)	(0.41)	(0.52)
AIC	4536.51	4770.07	3612.40	2616.25
Articles	3824	3824	3824	3824
Polls	402	402	402	402
Var (Intercept)	0.89	1.00	1.18	1.84

***p < 0.001, **p < 0.01, *p < 0.05

4 Alternative measure of change

We have used volatility between polls to measure change in our analysis as it fits with our case of multiparty competition. It accurately captures overall changes and can be extended to any other party system, both with more and with fewer parties regularly measured in the polls. However, there are alternative measures that would build on changes in the standing. Most certainly, this potential to measure overall changes through taking into account all party changes also makes it an unlikely candidate to being employed directly by journalists when evaluating the narrative potential and taking a decisions about selection.

Figure A4.1: The relationship between maximum change and previously studied outcomes



An accessible and intuitive heuristic for a change narrative is the greatest magnitude of change. Rather than summarizing it into one measure, for each poll, we looked at the maximum change, in absolute terms, a party registered compared to the previous poll from the same firm. This is directly apparent after looking at a poll in comparison to previous numbers, and allows for a more party centric coverage, i.e. the biggest winners and losers. We re-fitted all previous models, but instead of volatility, we used the maximum change (mean centred and standardized with two standard deviations). As we followed the same steps, these results should be directly comparable between the two operationalizations.

As displayed in Figure A4.1, we see very strong consistency in our results. This is not surprising, as the correlation between volatility and maximum change is 0.89. It is reassuring, as it shows that the empirical support for our theoretical model is not contingent on the specific operationalization, and that simple heuristics that might be more fitting for journalistic decision-making models work equally well.