



Not just the facts: an index for measuring the information density of political communication

Lauri Rapeli¹ · Sakari Nieminen² · Marko Mäkelä³

Accepted: 21 November 2020
© The Author(s) 2020

Abstract

Misinformation and biased opinion-formation plague contemporary politics. Fact-checking, the process of verifying accuracy of political claims, is now an expanding research area, but the methodology is underdeveloped. While the journalistic practice of fact-checking is by now well-established as an integral part of political news coverage, academic research requires more stringent methods than what journalists thus far have used. In order to advance the scientific study of fact-checking, we propose two variants of an index measuring the information density of verbal political communication. The main index combines three dimensions: (1) factual accuracy of political claims, (2) their relevance and (3) the magnitude of observed communication. In the article, we argue for the significance of each of these components. Depending on the research problem and data, the indices can be used for comparisons of political actors across different contexts such as countries or time points, or in non-comparative situations. Using examples, we demonstrate that the indices produce intuitive results.

Keywords Fact-checking · Index · Information density · Political communication · Political rhetoric

✉ Lauri Rapeli
lauri.rapeli@abo.fi

Sakari Nieminen
saanni@utu.fi

Marko Mäkelä
makela@utu.fi

¹ The Social Science Research Institute, Åbo Akademi University, Fänriksgatan 3 ASA A4, 20500 Åbo Finland, Finland

² Department of Philosophy, Contemporary History and Political Science, University of Turku, Turku, Finland

³ Department of Mathematics and Statistics, University of Turku, Turku, Finland

1 Introduction

Political communication is experiencing a period of rapid and continuing change (Krupnikov and Searles 2019). The declining quality of political information through news reporting has been identified as one of the key developments in the field. It is viewed as a particularly significant threat to the educative function of political communication through mass media (Van Aelst et al. 2017). Indeed, the emergence of phenomena such as ‘post-truth politics’, ‘fake news’ and ‘Twitter politics’, (e.g. Lazer et al. 2018) has given rise to an ever-growing literature dealing with various aspects of the (mis)use of facts in political speech. Scholars have looked at the role of misinformation in attitude formation, (e.g. Nyhan and Reifler 2010), the nature and prevalence of misperceptions, (e.g. Flynn, Nyhan, and Reifler 2017) and the journalistic practice of fact-checking, (e.g. Graves 2017). While there is a vivid scholarly debate on the epistemology of fact-checking (e.g. Graves 2017; Uscinski and Butler 2013), scholars have, however, mostly left it to the journalists to conduct empirical measurements of factual accuracy in contemporary political rhetoric.¹ Although fact-checking, i.e., the assessment of accuracy in public statements made by politicians, is by now a standard process for political media around the world,² academics have primarily focused on other issues regarding political (mis)information.

Consequently, the quantitative measurement of factual accuracy in verbal political communication has been primarily developed by and for journalists. The analytical depth of existing accuracy measures used by the media is adequate for journalistic purposes but from the point of view of academic research they are too one-dimensional. The existing methodology in verbal political communication has traditionally relied on various forms of content and discourse analysis designed to reveal the linguistic qualities of political rhetoric, but it cannot be used for a quantitative analysis of factual accuracy. Similarly, although many methodological approaches within political communication are quantitative, they cannot be applied to the measurement of factual accuracy. For example, the intersection between political communication and psychology has contributed quantitative measures of emotion and affect, to investigate the content of political rhetoric (e.g. Crigler and Just 2012). Moreover, the study of political communication has increasingly turned to automated analysis of political texts, to analyze politicians’ ideological stances using a combination of qualitative and quantitative methods (Sim et al. 2013; Laver, Benoit and Garry 2003). These techniques have led to important advances in addressing the essence of the study of political rhetoric – the strategies politicians use to construct persuasive arguments in political debate (Condor, Tileaga, and Billig 2013). However, these methods are also unsuitable for addressing factual accuracy.

Thus, the current scholarly interest across several scientific disciplines in the prevalence and use of facts in politics, is not supported by existing methodological tools. Fact-checking remains a primarily journalistic practice, and the methodology required for scientific inquiry

¹ However, see Meirick et al. (2018), where expert judgments were used to define the accuracy of political advertisements.

² The number of fact checkers in political journalism has grown markedly (Stencel 2017), along with the popularity of major US fact-checking websites (Mantzaris 2016). Between 2001 and 2012, mentions of fact-checking in US newspapers increased by over 900% and in TV and radio programs by over 2,000% (Amazeen 2013).

is still largely underdeveloped. To advance the methodology in the field, we propose two variants of an index measuring the information density of verbal political communication. The two versions are designed for different situations. The main index is more advanced and intended for situations where the context of the analysis is uniform enough to allow fair comparison between the units of analysis, using all the dimensions of the index. We also offer a simpler version, which can be used to compare actors from less uniform contexts.

The main index³ is three-dimensional, combining: (1) the factual accuracy of political claims, (2) the relevance of those claims and (3) the magnitude of the observed communication. Hence, the main index goes beyond the simple and straightforward measurement of factual accuracy to measure what we choose to call the information density of a political communication. The index looks simultaneously at both the accuracy and relevance of political claims, but also contextualizes that combination by including a component that measures the size of the communication in relation to the number of factual claims made by the actors included in the analysis. This last component allows users to distinguish between actors whose communication is mostly factual claims and actors who speak a lot but make only a few, if any, factual claims. We call the combination of these three components the 'information density', because this, as we argue, captures the broader idea of analyzing what is being claimed in politics and how the claims are made. The main index is intended specifically for analyses where it is important to take the amount of communication into account. For example, if one wants to analyze, which member of the German Bundestag has been the most informative in his/her parliamentary speeches during the last electoral term, our index offers a way to compress data about the accuracy, relevance and scope of his/her factual communication into one number that is comparable with corresponding numbers of other politicians in the same context.

The simpler version of the index is two-dimensional. It is otherwise identical with the main index, but it does not include the magnitude component. The simple index is designed for situations where units of analysis are not comparable in relation to the magnitude of the communication. This might be due to, e.g., different languages (that differ in average word and sentence lengths) or different communication platforms (as some platforms may have character limits). For instance, if one is interested whether President Donald Trump is more accurate and relevant in his State of the Union speeches than the Finnish President Sauli Niinistö is in his New Year's speeches, our main index is not suitable for measuring the information density, as those speeches are delivered in different languages, and thus the scope of communication component would skew the results. However, the simple index is created for situations like these, as it ignores the volume of communication, and thus makes it possible to compare politicians, even when they use different languages and communication platforms.

In the typical journalistic workflows, fact-checkers usually do not analyze all claims in particular context, but only the ones that they consider interesting and newsworthy. Thus, based on such practices, it is not possible to say anything about the overall truthfulness of politicians or make comparisons between them, although from a scholarly perspective these would be among the most interesting things to find out. Whereas journalistic fact-checking usually aims at producing news that consumers want to read, researchers are driven by accuracy concerns. Additionally, the traditional fact-checking methods only assess the factual correctness of a claim, but in order to compare the informativeness of politicians, observing only the accuracy is not enough. As we argue, at least the relevance of the claims made needs to be accounted for.

³ Throughout the article when we refer to 'the index' without explicit reference to a specific version of it, we mean this main index.

In our proposal, the values of the simple scale are set in a range between 1 and 4. All components of both indices are, however, manipulable according to the user's needs. By assigning weights, the researcher can specify the importance of each component. In what follows, we explain the key properties of the indices and use examples to demonstrate that the main scale produces intuitive results. We also show how weighting affects the results. Although discussing and meticulously defining concepts such as 'fact' and 'accuracy' is relevant for work in this field, we cannot engage deeply in those discussions within this study. Our ambition is strictly to offer analytical tools for quantitative, empirical research, without engaging deeply in the very important discussion on how, for example, the truth value of claims itself should be conceptualized. For contextualization, we begin by reviewing the journalistic process of fact-checking.

2 The current state of fact-checking methodology

The main purpose of fact-checking is to evaluate the truthfulness of political claims appearing in public, such as in politicians' speeches, interviews or social media messages. The accuracy of the claims is evaluated against various information sources, such as official statistics, documents and expert opinions. According to Graves (2017), a typical journalistic fact-check consists of five elements: (1) choosing claims to check, (2) contacting the speaker, (3) tracing false claims, (4) consulting experts and sources and (5) publishing the check as transparently as possible. The choice of claims involves considering the political significance of the claim and its newsworthiness, ascertaining that nonpartisan fact checkers have checked all sides of the matter and ensuring that the accuracy of the claim can actually be verified. By contacting the speaker, the fact checker gives a chance to offer an explanation. Then, by tracing a false claim, fact checkers try to discover the origin of an incorrect claim and to reconstruct its spread. When evaluating the claim, fact checkers rely on official data, and usually turn to experts for help in interpretation. To ensure transparency, the publication of the 'truth verdict' is accompanied by source documentation.

These five steps are not identical across different fact checkers. While most fact checkers select claims to check (based, e.g., on newsworthiness), some actors set explicit targets by defining beforehand how many claims they are going to check for each party, or whether they are going to check all claims in a particular debate or speech (Graves 2018). The Ukrainian organization StopFake, for example, is dedicated to countering misinformation that emanates (mostly) from the Russian media (Khaldarova and Pantti 2016).

While the process always ends in an accuracy evaluation of a claim, the accuracy scales used by the different fact checkers differ considerably. Some only use two categories: true or false, while others have much more nuanced scales. PolitiFact.com, which is one of the major US fact checkers, has a six-step 'Truth-O-Meter' scale. It ranges from 'true' to 'pants on fire'; the latter referring to claims that are outrageously false. The Washington Post's Fact Checker essentially only rates the degree of seriousness of mistakes in political claims by assigning Pinocchios. 'One Pinocchio' means 'Some shading of the facts. Selective telling of the truth. Some omissions and exaggerations, but no outright falsehoods.' The lowest category, four Pinocchios, is defined simply as a 'whopper'. A claim containing 'the truth, the whole truth and nothing but the truth' is awarded the Geppetto Checkmark (Kessler 2013). The scales therefore not only differ in terms of the number of categories, but also in the way they are constructed.

In addition to these differences in measurement, the reliability of the process itself has been called into question. Fact checkers have been accused of selecting claims to be checked based on what the journalists subjectively perceive as important and newsworthy (for different case selection principles, see, e.g., Adair and Drobnic Holan 2013; Kessler 2013; also Amazeen 2013). PolitiFact, for example, announces that ‘because we cannot possibly check all claims, we select the most newsworthy and significant ones’ (Adair and Drobnic Holan 2013). Critics claim that this may induce serious bias in fact-checking. If the checked claims are not a representative sample of the entire communication produced by some political actor, these actors may appear to be more accurate or inaccurate than they actually are. If case selection is not based on a balanced representation of different partisan views, we cannot rely on the findings to offer a reliable account of the accuracy of political actors in general (Uscinski and Butler 2013). Thus, the popular comparison of, e.g., Hillary Clinton’s and Donald Trump’s PolitiFact scorecards, does not necessarily provide us with a trustworthy picture of the accuracy of their overall communication, let alone their intentions to be honest. Fact checkers themselves acknowledge this and say that their purpose is to offer information about the accuracy of claims appearing in public, not to reveal who lies the most (see Amazeen 2013, 2015). Thus, journalistic fact-checks can offer valuable information about the truthfulness and background of individual claims, but they do not typically allow us to describe the accuracy of politicians in general, or to do meaningful comparisons between politicians. However, as will be argued below, if one wants to harness the potential of fact-checking as a political science method, the approach of focusing only on the most newsworthy claims is too narrow.

Fact-checking conducted by journalists is also arduous and requires a large, trained workforce. Attempting to provide a more efficient way to fact-check the vast universe of politics, the cutting edge of scholarship is developing automated fact-checking. This combines linguistics and computer science to create a fully computerized process of monitoring political communication, identifying verifiable claims and then verifying their correctness (see Babakar and Moy 2016 for a review).⁴ While automated fact-checking will undoubtedly be highly useful for many purposes, it cannot, at least in the foreseeable future, satisfy all the requirements of academic research. As we argue in the following section, merely checking the *accuracy* of a claim offers a too simplistic a view of the information content of political communication. Political issues vary in terms of importance and consequently, so do the truths and untruths told by politicians. Moreover, automated fact-checking is not particularly well-suited to comparing different types of political communication, and therefore lacks the ability to contextualize the use and misuse of facts in politics.

3 Measuring information density, not just accuracy

As noted in the previous section, fact-checking by journalists typically focuses on the accuracy of single claims made by individual politicians, often selected based on subjective assessments of newsworthiness. However, from a scientific viewpoint, such unsystematic selection is unacceptable, at least if the purpose is to analyze the overall accuracy of politicians’ communications. For scholars who are more concerned with comparative analysis, contextualization and the ‘big picture’, the accuracy and background of unsystematically selected individual claims is not particularly interesting (from an analytical point of view). Thus the first step in an attempt to use fact-checking as a political science method, is to

⁴ For a review of fact-checking literature in general, see Nieminen and Rapeli (2019).

define the boundaries of the communication that is included in the analysis, to establish the scope of the overall political communication from which claims will then be extracted. For example, instead of randomly choosing claims made by a politician, a researcher might want to examine all claims made by politicians from the two major US parties during a campaign in a congressional election, thus defining a specific temporal range for the communication by a given set of political actors which is to be fact-checked.

In addition, the existing journalistic accuracy scales are one-dimensional in a way that they do not assess the relevance of the checked claims. Without considering the relevance of the subject matter of a claim, all claims are treated as equal. But mistakes – and truths – come in different sizes. For instance, if an American politician confuses the names of Norwegian and Danish prime ministers, the mistake is, from an American perspective, relatively irrelevant compared to a mistake concerning, say, the content of the US Constitution or the development of key indicators of the national economy. For Norwegians and Danes, however, the relative significance of the same mistakes would probably be the opposite. Consequently, as we argue, the relevance of any claim must also be taken into account, if fact-checking (as a method) is to provide scientifically meaningful findings. As our example illustrates, by also measuring relevance it is possible to make comparisons across contexts where the same political issues may have very different significance. However, we are not claiming that journalistic fact-checkers' methods for assessing the truthfulness of claims (consulting experts, seeking official government data etc.) are somehow flawed and misplaced in general or that they should be abandoned. What we are arguing for, is that this accuracy evaluation process should be complemented with an assessment of the relevance of claims, if the full potential of fact-checking as a meaningful political science method is to be realized.

A closely related issue is the scope of the communication. Some politicians are (much) more talkative than others. While some politicians might produce several checkable claims in a one-minute speech, others might fail to make any checkable claims in the same period of time. If we do not take into account the magnitude of a politician's communication in relation to the number of factual claims made, actors with the same amount of communication but unequal numbers of checkable claims might seem equally informative. In a similar fashion as for the various formulas for estimating readability and the grade of difficulty in texts (see, e.g., Kayam (2018) for an overview), by including the size of a communication, we gain an understanding of how large or small the role of facts is in political communication. We can, for example, analyze the different rhetorical strategies of politicians in terms of the frequency of use of facts in their communication.

Since we argue here that a meaningful measurement of factual accuracy in political communication should go beyond the seemingly simple checking of factual accuracy, a broader term is needed to describe our indices. As we have already mentioned, we have chosen to call the combination of the three components—accuracy, relevance and the scope of the communication—the information density of political claims. Thus, by referring to 'information density', we wish to emphasize the need to contextualize the straightforward ratings of factual accuracy and offer a more advanced way of analyzing claims in political communication.

In addition, as the contexts for political communication are diverse, there might be situations where it is necessary, e.g. due to theoretical justifications, to emphasize one of these three components. For example, when analyzing statements made in a context where false claims could seriously harm the audience or some of its members (e.g. crisis communications in the case of a nationwide pandemic), it could be meaningful to stress the importance of accuracy over relevance. There might also be occasions when the researcher is

more interested in the political significance of the claims, for example, when analyzing statements politicians make in social media, and thus a possibility to raise relevance over truthfulness could be motivated. There might even be situations where the researcher wants to emphasize the scope of communication over truthfulness and relevance. For instance, in political debates, time is a scarce resource, and when one person speaks, it limits others' chances to do so. Thus, when analyzing the information density of politicians' debate performances there could be incentives to 'reward' those who focus on factual information and express themselves more sharply. Therefore, in our indices there exists an opportunity to weight these components of information density.

We now move on to introducing the two indices. We describe them simultaneously, and then recap and present ready-to-use formulas in the summary section.

4 The indices

The first component of the index, i.e., the factual accuracy of a political claim, is essentially a traditional fact-check, conducted to establish the truth-value of a claim. This part of the process of using the indices is essentially the same as the journalistic process: after identifying a claim for fact-checking, its accuracy is verified using the most reliable sources available.

A detailed account of how the users of our indices should determine the accuracy of the claims and define the categories of the scale they use, is beyond the scope of this study. We assume that the user is familiar with the journalistic practices of fact-checking, and that the chosen process for accuracy determination will resemble the procedures of journalistic fact-checkers, meaning that the accuracy ratings are based on evidence and nonpartisan sources, such as official government agencies and scientific studies, will be preferred. Transparency about the sources should be self-evident. As the datasets that the index is designed for are large, the practice of contacting the speaker, as some fact-checkers do (see Graves 2018), might be too labor-intensive. Given that the placing of claims in predetermined accuracy categories leaves the fact-checker with a considerably leeway when evaluating claims, e.g. when determining whether the claim could be considered completely accurate or only mostly accurate, inter-coder reliability testing is an essential part of academically motivated fact-checking. To decrease subjectivity, predefined margins of errors for numerical claims (e.g. how much a numerical claim is allowed to differ from the real number, and still be considered completely accurate) might be worth considering.⁵

The second component is the relevance measure, assessing the significance of the issue that the claim is addressing. While a thorough discussion of the exact nature of relevance is beyond the scope of this study (however, see e.g. Gorayska and Lindsey 1993 and Mizzaro 1997 for studies about relevance in general), it is nevertheless important to acknowledge some approaches on how relevance can be understood here. In the context of argumentation, Macagno (2019) has outlined a coding scheme for assessing the relevance of dialogue moves. The scheme divides the moves in a dialogue into irrelevant, weakly relevant and strongly relevant. The evaluation of 'dialogue moves' includes considerations about whether a move in a dialogue is coherent with the previous move(s). For example, an assertion is usually a relevant reaction to a question, while an order is not. It is also important

⁵ For example, Nieminen et. al. (2017) allowed a five percent margin of error for true claims in their fact-checking practice.

to consider whether a speaker's dialogical moves relate to the topic, whether the move contributes to the point of the dialogue and whether the move is based on acceptable and recognizable premises. An approach like this could very well be applicable in the context of our indices, when the density of political communication from a more argumentative perspective (i.e. whether politicians' claims are truthful and dialogically relevant) is to be assessed.

In the context of social media, some studies have assessed the relevance of users and followers. Cha et.al. (2010) have examined the influence of Twitter-users by comparing the number of followers, retweets and name mentions. Alsinet et.al. (2020), on the other hand, measure user relevance in Reddit debates by quantifying the number of comments, the influence of those comments, and whether the comments agree or disagree with the root comment. Even though these measures are not as such suitable for applying to our indices, the attention that a claim gets could be used as a basis for measuring the relevance in situations where the data is solely from social media platforms.⁶ A simple and straightforward way to do this could be, for example, to set quotas of likes or shares that each claim must reach in order to achieve a certain level of relevance. This is also an example of a situation where values for the relevance component in the indices can be obtained without a purely subjective assessment.

Sometimes relevance has been measured by the magnitude of the issue itself. For example, Quinn and Crean (2012) determine the relevance of small and medium sized service sector enterprises in different areas by using their percentage of the total workforce as an indication of significance (e.g. below 30% = irrelevant, 40% = important, 60% = highly relevant). Using such limit values could be another way to make relevance evaluations to be used in the indices. Examples from the realm of politics could include e.g. the amount of money that is connected to the claim's subject (for example, if the issue covers over 20% of nation's/city's/etc.'s budget, it is highly relevant, if 15%–20% it is relevant etc.) or the number of people that are allegedly affected (e.g. over 50% of the population could be 'highly relevant', 35%–50% could be 'relevant' and so forth).⁷

The above examples are intended as illustrations of how the question of 'relevance' has been dealt with in previous literature and how those solutions could be applied in the context of our indices. There is obviously no commonly accepted framework for ranking political issues in the terms of relevance, so the indices in this respect leave much leeway for the user, which highlights the importance of transparency in documenting the research process.

The third component is the scope of the overall communication that is included in the analysis. This part requires identifying the limits of the (verbal) communication that is incorporated into the analysis.

Let us denote the *accuracy* of a single claim by a and suppose that it has four possible integer values from 1 to 4 (with 4 meaning 'completely accurate' and 1 meaning 'completely inaccurate'). The user can freely decide on the range of values. Similarly, we denote the *relevance* of a claim by r , with the same four-point scale (4 meaning 'very relevant' and 1 meaning 'entirely irrelevant'), which can also be decided by the

⁶ The reasoning being that likes and shares could be seen to reflect what people consider important or interesting, and thus reflecting also relevance.

⁷ This idea was originally formed in conversations with professor Matti Wiberg.

Table 1 The influence of parameter q when q= 1

$a \setminus r$	1	2	3	4
1	1	1.41	1.73	2
2	1.41	2	2.44	2.83
3	1.73	2.44	3	3.46
4	2	2.83	3.46	4

Table 2 The influence of parameter q when q=2

$a \setminus r$	1	2	3	4
1	1	1.59	2.08	2.52
2	1.26	2	2.62	3.17
3	1.44	2.29	3	3.63
4	1.59	2.52	3.30	4

Table 3 The influence of parameter q when q= 10

$a \setminus r$	1	2	3	4
1	1	1.88	2.71	3.52
2	1.07	2	2.89	3.76
3	1.11	2.08	3	3.90
4	1.13	2.13	3.08	4

user. Arguably, in practically all situations, the researcher wishes to contrast higher accuracy and relevance values with lower values of both components. Therefore, in our index, both concepts are treated as having a parallel effect, so that we can multiply⁸ them in order to measure the information density of a single claim. The possibility of prioritizing accuracy or relevance is realized by adding positive weighting exponents p and q respectively. Thus, the *information density of a single claim* can be calculated as

$$s = (a^p r^q)^{\frac{1}{p+q}} \tag{1}$$

where the outer exponent $\frac{1}{p+q}$ scales the value of the information density to the same interval as the accuracy and the relevance. Notice that s is no longer an integer but is a real number between 1 and 4. Notice also that Eq. (1) can be rewritten in the form

$$s = (ar^{\frac{q}{1+\frac{q}{p}}})^{\frac{1}{1+\frac{q}{p}}} \tag{2}$$

This means, that we can always normalize to $p = 1$ and thus, for simplicity, it is enough to use only one exponent parameter q . In other words, instead of (1), we use the equation

$$s = (ar^q)^{\frac{1}{1+q}} \tag{3}$$

⁸ We could also use the simpler linear additive version of the index, where we sum the values instead of using multiplication. However, this nonlinear multiplicative version is more natural, since it penalizes extreme values. For example, $1 + 4 = 2 + 3$ but $1 \times 4 < 2 \times 3$.

For example, if we choose $q = 1$, the accuracy and the relevance are equally important, and the information density is then

$$s = \sqrt{ar} \quad (4)$$

In order to place more weight on the relevance, we can choose $q > 1$, e.g., $q = 2$, obtaining

$$s = \sqrt[3]{ar^2} \quad (5)$$

while the choice $q < 1$, e.g., $q = \frac{1}{2}$ emphasizes the accuracy, giving

$$s = \sqrt[3]{a^2r} \quad (6)$$

In Tables 1, 2, 3 we illustrate the influence of the parameter q by giving the information density values with all the possible combinations of the accuracy a and the relevance r . Table 1 corresponds the Eq. (4) meaning that the accuracy and the relevance are equally important. In Tables 2 and 3 we have put more importance to the relevance by choosing $q = 2$ and $q = 10$, respectively. Note that the diagonal values of the tables are not depending on the parameter q .

Suppose also, that instead of only one, there are several claims to be analyzed, as is typically the case. Let the total number of claims be n and the accuracy and the relevance of the i th claim be denoted by a_i and r_i respectively. Then, the natural way to evaluate the general information density is to calculate the average of the single items:

$$\bar{s} = \frac{1}{n} \sum_{i=1}^n (a_i r_i^q)^{\frac{1}{1+q}} \quad (7)$$

This is also the formula for the simple scale. Notice that since it is an average, \bar{s} is also a real number between 1 and 4.

The consideration of the third component, i.e., the scope of overall communication, is a rather more complicated task. This is due to the fact that the number of claims as a proportion of the used characters, depends strongly, for example, on the context and the language used. For instance, character limits in social media applications constrain the length of communications, but in many other types of communication such restrictions do not exist. More critically, languages differ from one another in terms of sentence length, and this also changes over time (Hammarström 2016; Bochkarev, Solovyev and Wichmann 2014; Sigurd et al. 2004). For this reason, we suppose that the average number of characters used per factual claim is a known constant, denoted by k . This constant k can be calculated by dividing the total number of characters used in the chosen context, (e.g., in a parliamentary debate) by the total number of factual claims made in the same context. If m is the total number of characters used by an individual politician and n is the number of factual claims made in the chosen context, then the ratio

$$\alpha = \frac{\frac{m}{n}}{\frac{k}{nk}} = \frac{m}{nk} \tag{8}$$

describes how many characters that politician uses to form a single claim, compared to the average.⁹ If $\alpha > 1$ there are fewer claims in the text than the average, and this should decrease the substance value (7), and vice versa.¹⁰ For this reason, we divide the general information density value \bar{s} by α^p , where p is again a positive weighting exponent. Thus, we obtain the *final information density index* in the form

$$S = \frac{\bar{s}}{\alpha^p} = \frac{n^{p-1}k^p}{m^p} \sum_{i=1}^n (a_i r_i^q)^{\frac{1}{1+q}} \tag{9}$$

Notice that if we choose $p = q = 1$, all the components, i.e., the accuracy, the relevance and the magnitude, are equally important, and the information density (9) reduces to

$$S = \frac{k}{m} \sum_{i=1}^n \sqrt{a_i r_i} \tag{10}$$

In order to place more weight on the magnitude, we can choose $p > 1$, e.g., $p = 2$, and then we obtain

$$S = \frac{nk^2}{m^2} \sum_{i=1}^n \sqrt{a_i r_i} \tag{11}$$

while the choice of $p < 1$, e.g., $p = \frac{1}{2}$, emphasizes the accuracy and relevance more, giving

$$S = \sqrt{\frac{k}{nm}} \sum_{i=1}^n \sqrt{a_i r_i} \tag{12}$$

⁹ The user of our main index should bear in mind that the scope of communication component penalizes speakers for all kinds of communication other than factual claims. However, there is plenty of relevant non-factual political communication, such as expressions of values. Thus, applying the main index to contexts where the factuality of communication is not necessarily a default (e.g. opinion pieces written by politicians), might not be appropriate. We will return to this issue in the discussion section.

¹⁰ Our decision to measure the scope of communication as the number of characters warrants explanation. Using the number of factual claims (proportioned to the number of all phrases) instead of characters would eliminate the problem of comparison between different languages. Such metrics have been used in analyses of political texts (see e.g. The Manifesto Project, where the coding units are quasi-sentences, Werner et.al 2015). This is certainly one potential direction for further development of our index, and something that we carefully considered when we were designing our index. However, we ended up using the number of characters, because it also takes into account how compact the communication is. The same content can usually be expressed in many different ways, either more concise (and maybe also clearer) or longer. Since we wish to address the information content of political communication, we designed the scope of communication component in a way that does not only take account only the number of claims that are made, but also the way how they are made. On the other hand, some facts are naturally longer in characters than others and therefore those who communicate about shorter and simpler facts get better characters per claim ratio than those who communicate about longer and more complicated issues. However, the character-based version might be more effective in penalizing the speaker for such practices as overcommunication (the speaker e.g. communicates something in more word-rich manner than necessary, in order to distract the audience, see e.g. Hansen 2015). In summation, there are arguments for both a claim-based and a character-based scope component, and after considering the issue we concluded that the character-based version might be more appropriate.

Notice that theoretically, S is no longer necessarily a number between one and four. As m increases, S tends to zero but is always positive. Extremes are theoretically possible, but do not occur in reality. On the other hand, since $1 \leq \bar{s} \leq 4$ and $m > n$, we have $\alpha > \frac{1}{k}$. This implies

$$\frac{\bar{s}}{\alpha^p} < 4k^p \quad (13)$$

and thus, we always have $0 < S < 4k^p$. However, in practice, α is often close to 1, meaning that the value of S is between 1 and 4.

5 Summary

Let us conclude by stating the ready-to-use formulas once more. The general form of the simple index, including only accuracy and relevance, is

$$\bar{s} = \frac{1}{n} \sum_{i=1}^n (a_i r_i^q)^{\frac{1}{1+q}} \quad (14)$$

where a_i is the accuracy of the i th claim ($a_i \in \{1, 2, 3, 4\}$), r_i is the relevance of the i th claim ($r_i \in \{1, 2, 3, 4\}$), n is the number of claims ($n > 0$) and q is the weighting exponent ($q > 0$).

If we choose $q = 1$, i.e., the accuracy and the relevance are equally important, (14) reduces to

$$\bar{s} = \frac{1}{n} \sum_{i=1}^n \sqrt{a_i r_i} \quad (15)$$

The general formula for the main index, which also includes the scope of communication, is

$$S = \frac{n^{p-1} k^p}{m^p} \sum_{i=1}^n (a_i r_i^q)^{\frac{1}{1+q}} \quad (16)$$

where in addition to the previously defined terms, k is the average number of characters used per factual claim ($k > 0$), m is the total number of characters used ($m > n$) and p is the weighting exponent ($p > 0$).

If we choose $p = q = 1$, i.e., the accuracy, the relevance and the magnitude are all equally important, (16) reduces to

$$S = \frac{k}{m} \sum_{i=1}^n \sqrt{a_i r_i} \quad (17)$$

Examples How should the indices be used in an empirical inquiry? In the following, we provide a few examples of different situations, to describe how the full index could be applied in scholarly work. The examples are analogous to real-life research designs, such as the fact-checks conducted by PolitiFact.com. Following the work logic of professional fact-checkers, our examples illustrate situations where the analysis examines verbal

communication by politicians. That communication includes claims that can be assigned a truth-value, and as we argue, a relevance value and a contextual component concerning the scope of communication.

Let us first consider the situation where a scholar is interested in comparing party manifestos in terms of their information density. In this case, the researcher should first choose which party platforms to analyze, thus defining the context of analysis. In the analysis, each individual manifesto could be treated as a separate analysis unit, but one could also combine, for example, several manifestos from one party into a single unit of analysis. By choosing a restricted context for the analysis, the researcher determines the scope of the communication that is included in the analysis. In the next phase, the researcher should identify checkable claims for each unit of analysis, (i.e., how many checkable claims each unit of analysis makes/contains) and assign accuracy and relevance values to each of them. The number of characters and factual claims should also be counted in each unit of analysis and these should then be summed in order to find the total number of characters and claims in the chosen context. When the total number of characters is divided by the total number of claims, this gives the value of k , which denotes the average number of characters used to produce one factual claim. This is then compared with the character-claim ratio for each unit in the analysis (see Eq. (8)), allowing the researcher to determine α for each analysis unit. The researcher is now ready to calculate the information density value for every analysis unit and thus reveal which party manifestos (or groups of manifestos or platforms) have the most accurate and relevant content.

Alternatively, a researcher could be interested in comparing the information density of a single politician over time; for example, across several parliamentary terms. In this case, each parliamentary term and the politician's speeches during each of them, form the units of analysis. The scholar should, again, identify checkable factual claims from each of the parliamentary terms and assign truth and relevance values to each of them. Characters used and checkable claims made, in each term, should be counted and summed to obtain the total number of characters and claims in the given context. Based on these values, the researcher can calculate k by dividing the total number of characters by the total number of claims. Using k , α can be calculated for each term (again, see Eq. (8)). The information density values can then be determined to investigate how informative the communication by the politician has been during the various parliamentary terms.

To consider one more possibility, sometimes a researcher wishes to compare actors whose communications have very different starting points. One might, for example, want to compare the information density of a politician's Twitter messages with blog posts by another politician, because the two politicians use those specific channels as their most important means of communication with their respective electorates. Because the nature of these two platforms is so different (Twitter has strict character restrictions while blogs usually do not), this may distort the information density value via the communication magnitude. Similarly, a researcher might want to compare communication by politicians from two different countries, e.g., parliamentary speeches concerning the implementation of a particular piece of European Union legislation. If the speakers use different languages, and the languages differ considerably in terms of average word and sentence length, this

Table 4 The information-content indices of example politicians' claims

Politician	No weighting	Accuracy emphasis (relevance weight $q = 0.5$)	Relevance emphasis (relevance weight $q = 2$)
1	5.37	5.37	5.37
2	2.68	3.38	2.13
3	2.68	2.13	3.38
4	1.34	1.34	1.34

Table 5 The impact of communication scope

Politician	No weighting	More weight to magnitude ($p = 2$)	Less weight to magnitude ($p = 0.5$)
5	3.58	3.20	3.78
6	1.53	0.59	2.48
7	1.34	0.60	2.01
8	2.68	2.40	2.84

reduces the main index value of the speakers who use the wordier language.¹¹ In these cases, the inclusion of the communication magnitude variable might not be meaningful, particularly if it is not weighted downward, but the simpler version of the index could still be used, focusing only on the accuracy and relevance of the claims. The simple index is also suitable if one only wants to analyze how one particular politician has performed in some particular context.

However, the parameter for communication magnitude can also be used to make comparisons across different contexts meaningful. Comparisons across time, political systems or different languages, for instance, can be made possible by using the magnitude parameter k , provided there is a universal reference point for the average communication magnitude. A universal reference point could, for example, be the average number of characters per claim across different countries or different measurement points in time. Using that average as a reference point for assessing the various individual measurements in the same data – countries or time points – puts them into context in situations where the different measurement points as such cannot be used for meaningful comparisons. In the absence of a proper universal reference point, the simpler index, however, offers a more practical solution.

In our example, we analyze data that contains 530,015 characters and 395 claims in all, the average number of characters per claim (k) thus being 1,342. In the first example, the data simulate comments made by four politicians during a parliamentary debate concerning a legislative proposal. The four politicians differ from one another in terms of accuracy and relevance in the claims they make in their comments, but the number of claims as well as the number of characters they use are identical. This example demonstrates how applying different weights to the accuracy or the relevance component, affects the results.

¹¹ Translation could, at least in principle, be the solution to this situation, but sometimes this is not feasible, (e.g., due to limited resources) or meaningful, (e.g., when something that affects the accuracy value of a claim could be lost in translation).

The example politicians in Table 4 have the following characteristics:

Politician 1: all claims are accurate and relevant (15 claims and 15,000 characters).

Politician 2: all claims are accurate but irrelevant (15 claims and 15,000 characters).

Politician 3: all claims are inaccurate but relevant (15 claims and 15,000 characters).

Politician 4: all claims inaccurate and irrelevant (15 claims and 15,000 characters).

As we can see from Table 4, weighting the accuracy or relevance does not affect the results if both components are identical for all claims (politicians 1 and 4). The situation is, however, different for politicians 2 and 3. When more emphasis was placed on accuracy by de-emphasizing relevance ($q=0.5$), the value of the information density index for politician 2 increased from 2.68 to 3.38, because his/her overall accuracy was higher than his/her overall relevance.¹² Correspondingly, the index value of politician 3 decreased from 2.68 to 2.13, because his/her overall relevance was higher than his/her overall accuracy. When the relevance was given more weight, the result was exactly the opposite: politician 3 received a clearly higher value, while the index value for politician 2 was reduced by the same amount.¹³

In the second example (Table 5), we demonstrate how the scope of the communication affects the results. Using the same data but different politicians, there are now two pairs of politicians. Within each pair the accuracy and relevance values are identical, but the magnitudes of their communication vary. Pair 1 could, for instance, illustrate a situation where the campaign messages of a leading politician are compared over time, with the first measurement point presenting political communication before the social media era and the second measurement point representing the present day. Pair 2 could instead be a comparison between two different types of communicators. Politician 7 is less inclined to make explicit claims with factual density, whereas politician 8 uses factual claims more often.

Pair 1:

Politician 5: all claims are accurate and relevant (10 claims and 15,000 characters).

Politician 6: all claims are accurate and relevant (10 claims and 35,000 characters).

Pair 2:

Politician 7: claims are mostly accurate and mostly relevant (5 claims and 15,000 characters).

Politician 8: claims are mostly accurate and mostly relevant (10 claims and 15,000 characters).

Firstly, it should be noted that all the politicians in Table 5 use more characters to make one factual claim than the average claim length in our example data. This is why each of these politicians has a lower index value when the magnitude is weighted upward and a higher value when the effect of magnitude is weighted downward. If they used fewer characters per claim than the average, the results would show an opposite trend.

¹² The weighting coefficient q manipulates the relevance component in the equation. Hence, assigning a weight of less than one increases the accuracy component in relation to relevance.

¹³ It is worth noting that the coefficients $q=0.5$ and $q=2$ are only used as examples, not as guidelines for future use of the index. From a technical standpoint, the applied weight can be any number.

In the first pair, politicians 5 and 6 make the same number of claims with the same accuracy and relevance values, but politician 6 uses more than twice the number of characters to make these claims. In other words, politician 6 produces much more communication than politician 5 in order to make one factual claim. This dramatically lowers his/her index value (1.53), compared to politician 5 (3.58). When the significance of communication magnitude is emphasized via the weight, the difference between the two politicians grows even larger because the number of characters per claim used by politician 5 is closer to the average of our data set than the corresponding number used by politician 6. Consequently, politician 5 is less affected by the magnitude weighting. However, when the effect of magnitude is lessened, politician 6 is more affected than politician 5, due to the fact that his/her amount of communication is greater.

For the second pair, politician 8 has a higher index value than politician 7 when there is no weighting, because he/she makes more claims with the same number of characters. This is also reflected in the weighted magnitude scenario, where politician 8 is less affected by the weight than politician 7. However, when the effect of communication magnitude is diminished via the weight (0.5), the difference between the two becomes significantly smaller.

To summarize, the above examples illustrate how manipulating the components of the index can considerably alter the index values, compared with the baseline situation where there is no weighting. However, we wish to stress that the weights we have applied in the examples only serve the purpose of demonstrating that there is a substantial impact on the results. They are not intended as a benchmark for future use. The use of any weighting should be based on theoretical grounds and the size of the weight is a matter for the researcher's judgement.

We conclude with a word of caution. If k is calculated from the overall number of characters and factual claims in the chosen context instead of using some universal reference point, this means that all comparisons between the analysis units should be made within that particular, predefined context. This means that one should not, for example, use the full index for a comparison of politicians in parliamentary debate A and parliamentary debate B, unless the overall numbers of characters and claims of both debates are added together or unless there is some universal reference point for k . If adding them together is not possible or meaningful, (e.g., due to different languages), but a comparison is nevertheless necessary, the simpler version of the index should be used instead.

6 Discussion

In this article, we have proposed two indices for measuring verbal political communication. The three-dimensional main index combines the factual accuracy of political claims, their relevance and the scope of overall political communication, to produce what we have termed 'the information density of political communication index'. We have presented an equation for a generic index that allows the user to manipulate the significance of the three dimensions by assigning weights to suit the researcher's analytical purposes. With the help of examples of typical analytical designs, we have demonstrated that the index produces intuitive results. Because the scope of communication component can sometimes be difficult to include in the analysis in meaningful way, (e.g., due language differences), we have also introduced a simpler version, which excludes this dimension.

Some limitations to our index warrant discussion. Firstly, establishing the exact scope of the political communication to be included in the analysis may be more difficult than it sounds. When considering the speed and volume of communication creation nowadays, it can be hard to find and collect all public messages that, for example, a particular politician makes, even if the time frame is precisely defined. The boundaries of relevant communication can, however, be easily defined in fixed contexts where the researcher is only concerned with a certain type of verbal communication, such as parliamentary debates from a certain period. Secondly, it is important to bear in mind that not all politically relevant communication is verbal. Non-verbal communication, such as gestures or body language, or paralanguage, such as voice pitch, rhythm and volume, can be a powerful means of asserting influence in politics and also of communicating 'information density'. Both of these aspects of communication remain outside the reach of our index, which is only capable of analyzing a body of text.

Thirdly, it is also important to bear in mind that not all politically relevant communication is of a 'factual' nature. Expressions of opinion, value statements and promises, for example, are all an essential part of political communication, but they have a negative effect on the values that our main index produces. The scope of the communication variable in the main index mathematically penalizes those who speak a great deal without making factual claims, compared to those who produce many factual assertions. If a politician communicates a great deal, but only a minor part of the communication contains factual claims, he/she performs worse according to the index than those who communicate the same amount but with more factual claims, all else being equal. This occurs even if the actual communication is politically relevant, despite not containing factual claims. Consequently, it is crucial for any researcher using our indices to understand that although the indices tap into what we have termed 'information density', other types of relevant political communication also exist.

Fourthly, the checkability of factual claims also remains outside the scope of our work in this study, but this is always an issue that any fact-checking project must deal with. The problem comes in many different shapes. Firstly, it is not always apparent how a claim can be identified for a fact-check to begin with. A rhetorical question or a sarcastic remark may contain an assertion that has the effect of a factual claim upon the listener, but it would be stretching the concept of fact-checking to consider such communication as a factual claim. It could also be that an accurate claim could be used in a misleading way. Secondly, even when communication contains a claim that can, in theory, be checked for accuracy, this may not always be possible in practice. Thirdly, the certainty with which the accuracy of claims can be verified, also varies. While some claims can be verified as either true or false beyond any reasonable doubt, others cannot. Perhaps the reliability of the fact-check itself ought to be incorporated as a fourth dimension, in the future development of the index. Finally, another limitation is that our index, like the practice of fact-checking in general, does not take the intention of the speaker into account. A false claim by a politician could be an innocent and honest mistake, but it could also be made with an intention to deceive. In our index, both cases are treated as equal, conceptually as well as mathematically. This is dictated by practice: revealing motivations is much more difficult than checking the information content of claims and may even be impossible (see Krebs and Jackson 2007). Our index is by no means a lie detector; it is a tool for assessing factual accuracy in political speech.

It is also useful to remember that our index simplifies political communication by placing claims in predetermined classes. As all claims must be subjectively evaluated for accuracy and relevance and placed in distinct categories, the subtle nuances of each

claim are lost and decisions about borderline cases between the different categories may cause instability in the findings. In fact, some fact checkers have even criticized the use of pre-existing truth-value categories, considering them inflexible and unscientific (see Graves 2018). Since our index requires the use of such categories, both for accuracy and relevance, it is imperative to be aware of the subjective element in the process of fact-checking (also Graves 2016: 210 ff.). The index values are only as reliable as the evaluations that produce them, which means that data transparency and intercoder reliability testing are key issues in the academically motivated process of fact-checking.

This process has taken a major step forward due to advances in automated fact-checking, which speeds up the time-consuming procedure of identifying and checking factual claims. We see our index as complementing and assisting scholarly work in this area. While automated processes can, optimally, provide scholars with quick and straightforward analyses of large quantities of data in terms of simply defined factual accuracy, our indices work offers a method for a more nuanced analysis of political communication. Our index aims at providing a numerical measure of information density for scholarly work that fact-checks political communication beyond the mere factual accuracy of statements by also considering the context in which claims are made and their significance as political claims.

Concluding on a more philosophical note, the significance of our index is based on the normative idea that political communication between politicians and citizens in a democratic polity should be truthful. While empirical research in the field does not have to be explicitly normative, the underlying idea is that more accuracy (or information density) is, generally speaking, better. However, in the context of international politics, for example, there might sometimes be good reasons for politicians to be deliberately inaccurate in their statements, even when addressing their own citizens. One might argue that in such cases, the potential benefits of outright lying might override the ethical principles against it (Mearsheimer 2011: 6–7). It could be that in some cases, a politician's duty would be to be untruthful, e.g., in cases of national security, and thus factually accurate statements about issues of high relevance might be a bad thing, although from the perspective of our index such statements would receive a high score. We therefore wish to emphasize that in academic inquiry, fact-checking does not in itself form any type of normative standard for assessing the conduct of politicians. However, with the help of our index, fact-checking can become a form of analysis for assessing the information density of verbal political communication in comparative settings.

Funding Open Access funding provided by Abo Akademi University (ABO). This research has been supported by the Kone Foundation, The Foundation for Municipal Development, The Strategic Research Council of the Academy of Finland (Grant No 312676) and The Abo Akademi University Foundation (grant to the FutuDem Centre of Excellence).

Compliance with ethical standards

Conflict of interest There are no conflict of interest to declare.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not

permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adair, B., Drobnic Holan, A.: The principles of PolitiFact, PunditFact and the Truth-O-Meter. PolitiFact. Retrieved from <http://www.politifact.com/truth-o-meter/article/2013/nov/01/principles-politifact-punditfact-and-truth-o-meter/> (2013).
- Alsinet, T., Argelich, J., Béjar, R., Martínez, S.: Measuring user relevance in online debates through an argumentative model. *Pattern Recogn. Lett.* (2020). <https://doi.org/10.1016/j.patrec.2020.02.008>
- Amazeen, M.A.: Making a difference? A critical assessment of fact checking in 2012. Research paper, New America Foundation (2013)
- Amazeen, M.A.: Revisiting the epistemology of fact checking. *Critical Review* (2015). <https://doi.org/10.1080/08913811.2014.993890>
- Babakar, M., Moy, W.: The State of automated fact-checking. Full Fact. Retrieved from https://fullfact.org/media/uploads/full_fact-the_state_of_automated_factchecking_aug_2016.pdf (2016)
- Bochkarev, V., Solovyev, V., Wichmann, S.: Universals versus historical contingencies in lexical evolution. *J. R. Soc. Interface* (2014). <https://doi.org/10.1098/rsif.2014.0841>
- Cha, M., Haddadi, H., Benevenuto, F., Gummadi, P.K.: Measuring user influence in Twitter: The million follower fallacy. In: *Proceedings of the Fourth International Conference on Weblogs and Social Media, ICWSM 2010, The AAAI Press*, 10–18. (2010)
- Condor, S., Tileaga, C., Billig, M.: Political rhetoric. In: Huddy, L., Sears, D.O., Levy, J.S. (eds.) *Oxford handbook of political psychology*. Oxford University Press, Oxford (2013)
- Crigler, A.N.: Just. M. R.: Measuring affect, emotion and mood in political communication. In: Semetko, H.A., Scammell, M. (eds.) *The Sage handbook of political communication*. Sage, Thousand Oaks (2012)
- Flynn, D.J., Nyhan, B., Reifler, J.: The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics. *Adv. Political Psychol.* (2017). <https://doi.org/10.1111/pops.12394>
- Gorayska, B., Lindsay, R.: The roots of relevance. *J. Pragmat.* (1993). [https://doi.org/10.1016/0378-2166\(93\)90091-3](https://doi.org/10.1016/0378-2166(93)90091-3)
- Graves, L.: *Deciding what's true: The rise of political fact checking in American journalism*. Columbia University Press, New York (2016)
- Graves, L.: Anatomy of a fact check: Objective practice and the contested epistemology of fact checking. *Commun. C. Crit.* (2017). <https://doi.org/10.1111/cccr.12163>
- Graves, L.: Boundaries not drawn: Mapping the institutional roots of the global fact checking movement. *Journal. Stud.* (2018). <https://doi.org/10.1080/1461670X.2016.1196602>
- Hammarström, H.: Linguistic diversity and language evolution. *J. Lang. Evol.* (2016). <https://doi.org/10.1093/jole/lzw002>
- Hansson, S.: Calculated overcommunication: Strategic uses of prolixity, irrelevance, and repetition in administrative language. *J. Pragmat.* (2015). <https://doi.org/10.1016/j.pragma.2015.05.014>
- Kayam, O.: The readability and simplicity of Donald Trump's language. *Political Stud. Rev.* (2018). <https://doi.org/10.1177/1478929917706844>
- Kessler, G.: About the Fact Checker. *The Washington Post*. Retrieved from https://www.washingtonpost.com/news/fact-checker/about-the-fact-checker/?utm_term=.ad867f47b828 (2013, September 11)
- Khaldarova, I., Pantti, M.: Fake news: The narrative battle over the Ukrainian conflict. *Journal. Pract.* (2016). <https://doi.org/10.1080/17512786.2016.1163237>
- Krebs, R.R., Jackson, P.T.: Twisting tongues and twisting arms: The power of political rhetoric. *Eur. J. Int. Relat.* (2007). <https://doi.org/10.1177/1354066107074284>
- Krupnikov, Y., Searles, K.: New approaches to method and measurement in the study of political communication effects. *Political Commun.* (2019). <https://doi.org/10.1080/10584609.2018.1526239>
- Laver, M., Benoit, K., Garry, J.: Extracting policy positions from political texts using words as data. *Am. Political Sci. Rev.* (2003). <https://doi.org/10.1017/S0003055403000698>
- Lazer, D.M.J., Baum, M.A., Benkler, Y., Berinsky, A.J., Greenhill, K.M., Menczer, F., Zittrain, J.L.: The science of fake news. *Science* (2018). <https://doi.org/10.1126/science.aao2998>
- Macagno, F.: Coding relevance. *Learn C. Soc. Interact.* (2019). <https://doi.org/10.1016/j.lcsi.2019.100349>

- Mantzlarlis, A.: Fact checking: Does anyone even care? The Poynter Institute. Retrieved from <https://www.poynter.org/news/fact-checking-does-anyone-even-care> (2016, January 8)
- Mearsheimer, J.: Why leaders lie: The truth about lying in international politics. Oxford University Press, Oxford (2011)
- Meirick, P.C., Nisbett, G.S., Harvell-Bowman, L.A., Harrison, K.J., Jefferson, M.D., Kim, T., Pfau, M.W.: To tell the truth: Ad watch coverage, ad tone, and the accuracy of political advertising. *Political Commun.* (2018). <https://doi.org/10.1080/10584609.2017.1414089>
- Mizzaro, S.: Relevance: The whole history. *J. Am. Soc. Inform. Sci.* (1997). [https://doi.org/10.1002/\(SICI\)1097-4571\(199709\)48:9%3c810::AID-ASI6%3e3.0.CO2-U](https://doi.org/10.1002/(SICI)1097-4571(199709)48:9%3c810::AID-ASI6%3e3.0.CO2-U)
- Nieminen, S., Raiskila, M., Wiberg, M.: Faktat politiikassa [Facts in politics]. Helsinki, The Foundation for Municipal Development (2017)
- Nieminen, S., Rapeli, L.: Fighting misperceptions and doubting journalists' objectivity: A review of fact checking literature. *Political Stud. Rev.* (2019). <https://doi.org/10.1177/1478929918786852>
- Nyhan, B., Reifler, J.: When corrections fail: The persistence of political misperceptions. *Polit. Behav.* (2010). <https://doi.org/10.1007/s11109-010-9112-2>
- Quinn, M., Crean, A.: Helping small firms. *Account. Irel.* **44**(2), 35–37 (2012)
- Sigurd, B., Eeg-Olofsson, M., Van Weijer, J.: Word length, sentence length and frequency – Zipf revisited. *Studia Linguistica* (2004). <https://doi.org/10.1111/j.0039-3193.2004.00109.x>
- Sim, Y., Acree, B. D. L., Gross, J. H., Smith, N. A.: Measuring ideological proportions in political speeches. In: Yarowsky, D., Baldwin, T., Korhonen, A., Livescu, K., Bethard, S. (eds.) *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 91–101. Stroudsburg: ACL (2013)
- Stencel, M.: International fact checking gains ground, Duke census finds. Duke Reporters' Lab. Retrieved from <https://reporterslab.org/international-fact-checking-gains-ground/> (2017, February 28)
- Uscinski, J.E., Butler, R.W.: The epistemology of fact checking. *Crit. Rev.* (2013). <https://doi.org/10.1080/08913811.2013.843872>
- Van Aelst, P., Strömbäck, J., Aalberg, T., Esser, F., de Vreese, C., Matthes, J., Hopmann, D., Salgado, S., Hubé, N., Stepińska, A., Papathanassopoulos, S., Berganza, R., Legnante, G., Reinemann, C., Sheafer, T., Stayner, J.: Political communication in a high-choice media environment: a challenge for democracy? *Ann. Int. Commun. Assoc.* (2017). <https://doi.org/10.1080/23808985.2017.1288551>
- Werner, A., Lacewell, O., Volkens A.: *Manifesto Coding Instructions* (5th revised edition). <https://manifesto-project.wzb.eu/information/documents/handbooks> (2015, February)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.