



ARTICLE



<https://doi.org/10.1057/s41599-022-01174-9>

OPEN

# The fingerprints of misinformation: how deceptive content differs from reliable sources in terms of cognitive effort and appeal to emotions

Carlos Carrasco-Farré  <sup>1</sup>✉

Not all misinformation is created equal. It can adopt many different forms like conspiracy theories, fake news, junk science, or rumors among others. However, most of the existing research does not account for these differences. This paper explores the characteristics of misinformation content compared to factual news—the “fingerprints of misinformation”—using 92,112 news articles classified into several categories: clickbait, conspiracy theories, fake news, hate speech, junk science, and rumors. These misinformation categories are compared with factual news measuring the cognitive effort needed to process the content (grammar and lexical complexity) and its emotional evocation (sentiment analysis and appeal to morality). The results show that misinformation, on average, is easier to process in terms of cognitive effort (3% easier to read and 15% less lexically diverse) and more emotional (10 times more relying on negative sentiment and 37% more appealing to morality). This paper is a call for more fine-grained research since these results indicate that we should not treat all misinformation equally since there are significant differences among misinformation categories that are not considered in previous studies.

<sup>1</sup>Universitat Ramon Llull—ESADE Business School, Barcelona, Spain. ✉email: [carlos.carrasco@esade.edu](mailto:carlos.carrasco@esade.edu)

## Introduction

How can we mitigate the spread of disinformation and misinformation? This is one of the current burning questions in social, political and media circles across the world (Kietzmann et al., 2020). The answer is complicated because the existing evidence shows that the prominence of deceptive content is driven by three factors: volume, breadth, and speed. While the access to information has been dramatically increasing since the advent of internet and social networks, the volume of misleading and deceptive content is also on the rise (Allcott et al., 2018); that is the volume problem. Also, deceptive content can adopt different forms. Misinformation can appear as rumors, clickbait or junk science (trying to maximize visitors to a webpage or selling “miraculous” products) or in the form of fake news or conspiracy theories (false information spread deliberately to affect political or social institutions) (Scheufele & Krause, 2019); that is the breadth problem. Finally, misinformation spreads six times faster than factual information (Vosoughi et al., 2018), that is the speed challenge. These three factors contribute to making the tackling of both disinformation and misinformation one of the biggest problems in our times.

Some of the prominent social media platforms have intended to stop the proliferation of misinformation with different features like relying on user’s reporting mechanisms (Chan et al., 2017; Lewandowsky et al., 2012) or using fact-checkers to analyze content that already went viral through the network (Chung & Kim, 2021; Tambuscio et al., 2018; Tambuscio et al., 2015). However, both approaches have some shortcomings. The assessment of third-party fact-checkers or user’s reports is sometimes a slow process and misinformation often persists after being exposed to corrective messages (Chan et al., 2017). Therefore, a purely human-centered solution is ineffective because misinformation is created in more quantity (volume), in more forms (breadth) and faster (speed) than the human ability to fact-check everything that is being shared in a given platform. To address this issue, I propose to explore what I call “the fingerprints of disinformation”. That is, how factual news differ from different types of misinformation in terms of (1) evocation to emotions (sentiment analysis—positive, neutral, negative—and appeal to moral language as a challenge to social identity), and (2) cognitive effort needed to process the content (both in terms of grammatical features—readability—and lexical features—perplexity).

Regarding cognitive effort, extant research in the human cognition and behavioral sciences can be leveraged to identify misinformation through quantitative measures. For example, the information manipulation theory (McCornack et al., 2014) proposes that misinformation is expressed differently in terms of writing style. The main intuition is that misinformation creators have a different writing style seeking to maximize reading, sharing and, in general, maximizing virality. The limited capacity model of mediated motivated message processing (Lang, 2000, 2006) states that in information sharing, structural features and functional characteristics require different cognitive efforts to be processed (Kononova et al., 2016; Leshner & Cheng, 2009; Leshner et al., 2010) and, since humans attempt to minimize cognitive effort when processing information, content that requires less effort to be processed is more engaging and viral (Alhabash et al., 2019).

Moreover, another fingerprint of misinformation is its reliance on emotions (Bakir & McStay, 2018; Kramer et al., 2014; Martel et al., 2019; Taddicken & Wolff, 2020). In general, content that evokes high-arousal emotions is more viral (Berger, 2011; Berger & Milkman, 2009; Berger & Milkman, 2013; Goel et al., 2015; Milkman & Berger, 2014), which explains why social networks are a source of massive-scale emotional contagion (Fowler &

Christakis, 2009; Kramer et al., 2014; Rosenquist et al., 2011). One of the main reasons proposed to explain this behavior is the dual-process theory of judgment stating that emotional thinking (in contrast to a more analytical thinking) hinders good judgment (Evans, 2003; Stanovich, 2005). Indeed, there is experimental evidence that engaging in analytic thinking reduces the propensity to share fake news (Bago et al., 2020; Gordon Pennycook et al., 2015; Gordon Pennycook & Rand, 2019). For example, encouraging people to think analytically, in contrast to emotionally, decreases likelihood of “liking” or sharing fake news (Effron & Raj, 2020). On the contrary, reliance on emotion is associated with misinformation sharing (Weeks, 2015) or believing in conspiracy theories (Garrett & Weeks, 2017; Martel et al., 2019). Similarly, tweets with negative content are retweeted more rapidly and frequently than positive or neutral tweets (Tsugawa & Ohsaki, 2017) and sentiment rather than the actual information content predicts engagement (Matalon et al., 2021). Moreover, beyond sentiment, there are other measures for evocation to emotions: morality. Although there is little research on how moral content contributes to online virality (Rathje et al., 2021), the mechanism is grounded in social identity theory (Tajfel & Turner, 1979) and self-categorization theory (Turner et al., 1987) and lies in the idea that group identities are hyper salient on social media (Brady et al., 2017) because they act as a form of self-conscious identity representation (Kraft et al., 2020; van Dijck, 2013). Therefore, content that appeals to what individuals think is moral (or not moral), is likely to be more viral (Brady et al., 2020), especially in conjunction with negative emotions that challenge their social identity (Brady et al., 2017; Horberget al., 2011). In contrast, reliable sources are forced to follow the principle of objectivity typical of good journalistic practice (Neuman et al., 1992).

This unprecedented fingerprint of misinformation provides evidence that content features differ significantly between factual news and different types of misinformation and therefore can facilitate early detection, automation, and the use of intelligent techniques to support fact-checking and other mitigation actions. More specifically, the novelty and benefits of the paper are fourfold:

1. *Volume benefits: A solution must be scalable.* This proposal is a highly scalable technique that relies on psychological theories and Natural Language Processing methods to discern between factual news and misinformation types of content. Summarizing, I propose that misinformation has different complexity levels (in terms of lexical and grammatical features) that require different levels of cognitive effort and that misinformation evokes high-arousal emotions and a higher appeal to moral values. Using the mentioned variables, we can quickly classify every content shared in social networks at the exact moment when it is posted with little human intervention, helping to mitigate the volume challenge.
2. *Breadth benefits: Not all disinformation is created equal.* While extant research has been dedicated to differentiating between factual and fake news aiming at binary classification (Choudhary & Arora, 2021; de Souza et al., 2020), there is not much evidence regarding rumors, conspiracy theories, junk science or hate speech altogether. In this paper, I propose a model that can differentiate between 7 different categories of content: clickbait, conspiracy theories, fake news, hate speech, junk science, rumors, and finally, factual sources. To my knowledge, this is the paper with higher number of categories and higher number of news articles.

3. *Time benefits: Detecting misinformation before it is too late with on-spot interception.* Usually, fact-checkers and platform create lists and rankings of content (usually URLs) that are viral to assess its veracity. However, this procedure has a problem: the content is fact-checked once it has gone viral. In this paper I propose a method that is independent of network behavior, information cascades or their virality. Therefore, it allows to identify misinformation before it spreads through the network.
4. *Explainability: Avoiding inmates running the asylum.* In contrast to previous approaches (Hakak et al., 2021; Mahir et al., 2019; Manzoor et al., 2019), I provide an explainable model that is well-justified and grounded in common methods. In other words, I employ a set of mechanics that are easily explainable in human terms. This is important because this type of model have been rarely available (Miller et al., 2017). This model allows not only researchers but also practitioners and non-technical audiences to understand, and potentially adopt the model.

In particular, I analyze 92,112 news articles with a median of 461 words per article (with a range of 201 to 1961 words and 58,087,516 total words) to uncover the features that differentiate factual news and six types of misinformation categories using a multinomial logistic regression (see Methods). For this, I assess the importance of quantitative features grounded in psychological and behavioral sciences through an operationalization based on Natural Language Processing. Specifically, I assess the readability, perplexity, evocation to emotions, and appeal to morality of the 92,112 articles. Among others, these results show that fake news and conspiracy theories are, on average, 8% simpler in terms of readability and 18% simpler in terms of perplexity, and 18% more reliant on negative emotions and 45% more appealing to morality than factual news.

## Methods

**Data.** In order to carry out the analysis I use the Fake News Corpus (Szapkowski, 2018), comprised of 9.4 million news items extracted from 194 webpages. Beyond the title and content for each item, the corpus also categorizes each new into one of the following categories: clickbait, conspiracy theories, fake news, hate speech, junk science, factual sources, and rumors. The category for each website is extracted from the OpenSources project (Zimdars, 2017). In particular, the definitions for each category are:

*Clickbait.* Sources that provide generally credible content, but use exaggerated, misleading, or questionable headlines, social media descriptions, and/or images.

*Conspiracy Theory.* Sources that are well-known promoters of kooky conspiracy theories.

*Fake News.* Sources that entirely fabricate information, disseminate deceptive content, or grossly distort actual news reports.

*Hate News.* Sources that actively promote racism, misogyny, homophobia, and other forms of discrimination.

*Junk Science.* Sources that promote pseudoscience, metaphysics, naturalistic fallacies, and other scientifically dubious claims.

*Reliable.* Sources that circulate news and information in a manner consistent with traditional and ethical practices in journalism.

*Rumor.* Sources that traffic in rumors, gossip, innuendo, and unverified claims.

The labeling of each website was done through crowdsourcing following the instructions as follows (Zimdars, 2017):

Step 1: Domain/Title analysis. Here the crowdsourced participants look for suspicious domains/titles like “com.co”.

Step 2: About Us Analysis. The crowdsourced participants are asked to Google every domain and person named in the About Us section of the website or whether it has a Wikipedia page with citations.

Step 3: Source Analysis. If the article mentions an article or source, participants are asked to directly check the study or any cited primary source. Then, they asked to assess if the article accurately reflects the actual content.

Step 4: Writing Style Analysis. Participants are asked to check if there is a lack of style guide, a frequent use of caps, or any other hyperbolic word choices.

Step 5: Esthetic Analysis. Similar to the previous step, but focusing in the esthetics of the website, including photo-shopped images.

Step 6: Social Media Analysis. Participants are asked to analyze the official social media users associated with each website to check if they are using any of the strategies listed above.

Here, it is important to note that I follow a similar approach as previous studies to classify sources (Broniatowski et al., 2022; Cinelli et al., 2020; Cox et al., 2020; Lazer et al., 2018; Singh et al., 2020). For example, Bovet and Makse (2019) use Media Bias Fact Check to classify tweets according to their crowdsourced classification instead of manually classifying tweet by tweet. This approach has two advantages. First, that it is scalable (Broniatowski et al., 2022) and, secondly, misinformation intent is better captured at the source level than at the article level (Grinberg et al., 2019). Being aware of the potential limitations of this method, the approach offers a benefit: being able to tackle the breadth problem.

Moreover, the fact that I include several misinformation categories is important because most of the existing research has a strong emphasis on distinguishing between fake news and factual news (de Souza et al., 2020; Helmstetter & Paulheim, 2018; Masciari et al., 2020; Zervopoulos et al., 2020). However, not all misinformation is created equal. In general, it is accepted that there are several categories of misinformation delimited by its authenticity and intent. Authenticity is related to the possibility of fact-checking the veracity of the content (Appelman & Sundar, 2016). For example, a statistical fact is easily checkable. However, conspiracy theories are non-factual, meaning we are not able to fact-check their veracity. On the other hand, intent can vary between mislead the audience (fake news or conspiracy theories), attract website traffic (clickbait) or undefined intent (rumors) (Tables 1 and 2).

In the Fake News Corpus, each website is categorized among one of the options and all their articles have the corresponding category. From there, I extracted 30,000 random articles from each category, generating a dataset of 210,000 misinformation articles. For factual news, I used Factiva to download articles from *The New York Times*, *The Wall Street Journal* and *The Guardian*. This resulted in 3177 articles. In total, the database consists of 213,177 articles. I filtered the articles with less than 200 words and those with more than 2000 words, ending up with a database of 147,550 articles. After calculating all the measures (readability scores, perplexity, appeal to morality and sentiment analysis), I deleted all the outliers (lower bound quantile = 0.025 and upper bound quantile = 0.975). This resulted in the final dataset consisting of 92,112 articles with the following distribution by type: clickbait (12,955 articles), conspiracy theories (15,493 articles), fake news (16,158 articles), hate speech (15,353 articles), junk science (16,252 articles), factual news (1743 articles) and rumors (14,158 articles). The 197 websites hosting the 92,112 articles are:

**Table 1 List of sources and their corresponding category.**

ID	URL	TYPE	ID	URL	TYPE	ID	URL	TYPE	ID	URL	TYPE	ID	URL	TYPE
1	bipartisanreport.com	Clickbait	51	success-street.com	Fake News	101	healthimpactnews.com	Junk	151	counterpysops.com	Junk	151	counterpysops.com	Conspiracy
2	blacklistednews.com	Clickbait	52	gopthedailydose.com	Fake News	102	realfarmacy.com	Science	152	nowtheendbegins.com	Junk	152	nowtheendbegins.com	Conspiracy
3	breaking911.com	Clickbait	53	itaglive.com	Fake News	103	ewao.com	Science	153	infowars.com	Junk	153	infowars.com	Conspiracy
4	lifesitenews.com	Clickbait	54	thenewyorkevening.com	Fake News	104	naturalblaze.com	Science	154	thepoliticalinsider.com	Junk	154	thepoliticalinsider.com	Conspiracy
5	yournewswire.com	Clickbait	55	learnprogress.org	Fake News	105	foodbabe.com	Science	155	politicalblindspot.com	Junk	155	politicalblindspot.com	Conspiracy
6	addictinginfo.org	Clickbait	56	now8news.com	Fake News	106	responsibletechnology.org	Science	156	oilgeopolitics.net	Junk	156	oilgeopolitics.net	Conspiracy
7	politicususa.com	Clickbait	57	freedomdaily.com	Fake News	107	in5d.com	Science	157	ufoholic.com	Junk	157	ufoholic.com	Conspiracy
8	lifezette.com	Clickbait	58	rhotv.com	Fake News	108	whydonyoutrythis.com	Science	158	informationclearinghouse.info	Junk	158	informationclearinghouse.info	Conspiracy
9	twitchy.com	Clickbait	59	downtrend.com	Fake News	109	collective-evolution.com	Science	159	worldtruth.tv	Junk	159	worldtruth.tv	Conspiracy
10	bluenationreview.com	Clickbait	60	dailybuzzlive.com	Fake News	110	galacticconnection.com	Science	160	assassinationscience.com	Junk	160	assassinationscience.com	Conspiracy
11	occupydemocrats.com	Clickbait	61	thebigmiddle.com	Fake News	111	wakingtimes.com	Junk	161	secretsofthefed.com	Junk	161	secretsofthefed.com	Conspiracy
12	liberalamerica.org	Clickbait	62	ushealthyadvisor.com	Fake News	112	revolutions2040.com	Science	162	davidwolfe.com	Junk	162	davidwolfe.com	Conspiracy
13	other98.com	Clickbait	63	usa-television.com	Fake News	113	healthnutnews.com	Science	163	hangthebankers.com	Junk	163	hangthebankers.com	Conspiracy
14	threepercenternation.com	Fake News	64	healthycareandbeauty.com	Fake News	114	healthy-holistic-living.com	Science	164	corbetteport.com	Junk	164	corbetteport.com	Conspiracy
15	coed.com	Fake News	65	americanoverlook.com	Fake News	115	dineal.com	Science	165	eyeopening.info	Junk	165	eyeopening.info	Conspiracy
16	therightscoop.com	Fake News	66	smag31.com	Fake News	116	geoengineeringwatch.org	Science	166	govtslaves.info	Junk	166	govtslaves.info	Conspiracy
17	thefreepatriot.org	Fake News	67	yesimright.com	Fake News	117	fellowshipoftheminds.com	Conspiracy	167	whatreallyhappened.com	Conspiracy	167	whatreallyhappened.com	Conspiracy
18	newsto.com	Fake News	68	redcountry.us	Fake News	118	thedailysheep.com	Conspiracy	168	wikispoons.com	Conspiracy	168	wikispoons.com	Conspiracy
19	clashdaily.com	Fake News	69	usadosenews.com	Fake News	119	abovetopsecret.com	Conspiracy	169	theeconomiccollapseblog.com	Conspiracy	169	theeconomiccollapseblog.com	Conspiracy
20	usasupreme.com	Fake News	70	thenet24h.com	Fake News	120	familysecuritymatters.org	Conspiracy	170	henrymakow.com	Conspiracy	170	henrymakow.com	Conspiracy
21	weeklyworldnews.com	Fake News	71	flashnews.com	Fake News	121	conservativefocus.com	Conspiracy	171	whatdoesitmean.com	Conspiracy	171	whatdoesitmean.com	Conspiracy
22	prntly.com	Fake News	72	channel18news.com	Fake News	122	freedomoutpost.com	Conspiracy	172	whowhatwhy.org	Conspiracy	172	whowhatwhy.org	Conspiracy
23	thetruthdivision.com	Fake News	73	usinfonews.com	Fake News	123	greenvillepost.com	Conspiracy	173	zoofeed.com	Conspiracy	173	zoofeed.com	Conspiracy
24	theinternetpost.net	Fake News	74	news4kila.com	Fake News	124	dciothesline.com	Conspiracy	174	nodisinfo.com	Conspiracy	174	nodisinfo.com	Conspiracy
25	dailyurge.com	Fake News	75	onepoliticalplaza.com	Fake News	125	americanfreepress.net	Conspiracy	175	angrypatriotmovement.com	Conspiracy	175	angrypatriotmovement.com	Conspiracy
26	70news.worldpress.com	Fake News	76	empireherald.com	Fake News	126	newstarget.com	Conspiracy	176	libertytalk.fm	Conspiracy	176	libertytalk.fm	Conspiracy
27	newswithviews.com	Fake News	77	enlilve.com	Fake News	127	zerohedge.com	Conspiracy	177	fprnradio.com	Conspiracy	177	fprnradio.com	Conspiracy
28	teaparty.org	Fake News	78	thewashingtonpress.com	Fake News	128	canadafreepress.com	Conspiracy	178	sheepkillers.com	Conspiracy	178	sheepkillers.com	Conspiracy
29	president45donaldrump.com	Fake News	79	openmagazines.com	Fake News	129	awarenessact.com	Conspiracy	179	82.221.129.208	Conspiracy	179	82.221.129.208	Conspiracy
30	rickwells.us	Fake News	80	donaldrumpnews.co	Fake News	130	21stcenturywire.com	Conspiracy	180	blackgenocide.org	Conspiracy	180	blackgenocide.org	Conspiracy
31	newsfrompolitics.com	Fake News	81	uspoln.com	Fake News	131	activistpost.com	Conspiracy	181	amren.com	Conspiracy	181	amren.com	Hate

**Table 1 (continued)**

ID	URL	TYPE	ID	URL	TYPE	ID	URL	TYPE	ID	URL	TYPE
32	dcgazette.com	Fake News	82	goneleft.com	Fake News	132	infiniteunknknown.net	Conspiracy	182	returnofkings.com	Hate
33	enduringvision.com	Fake News	83	thelastgreatstand.com	Fake News	133	humansarefree.com	Conspiracy	183	themuslimissue.wordpress.com	Hate
34	onlineconservativepress.com	Fake News	84	newsmagazine.com	Fake News	134	theeventchronicle.com	Conspiracy	184	barnesreview.org	Hate
35	thecommonsenseshow.com	Fake News	85	worldpoliticsnow.com	Fake News	135	fromthetrenchesworldreport.com	Conspiracy	185	drrichswier.com	Hate
36	realnewsrightnow.com	Fake News	86	metropolitanworlds.com	Fake News	136	prisonplanet.com	Conspiracy	186	nationalvanguard.org	Hate
37	conservativedailypost.com	Fake News	87	civictribune.com	Fake News	137	pamelageller.com	Conspiracy	187	ihf.org	Hate
38	anonjckloy.tk	Fake News	88	stormcloudsgathering.com	Fake News	138	neonnettle.com	Conspiracy	188	therightstuff.biz	Hate
39	vigilantcitizen.com	Fake News	89	usanewsflash.com	Fake News	139	shoebat.com	Conspiracy	189	barenakedislam.com	Hate
40	empirenews.net	Fake News	90	proudcons.com	Fake News	140	endotheamericandream.com	Conspiracy	190	americanborderpatrol.com	Hate
41	usadailytime.com	Fake News	91	americantoday.news	Rumor	141	educate-yourself.org	Conspiracy	191	davidduke.com	Hate
42	dailyheadlines.com	Fake News	92	express.co.uk	Rumor	142	thelibertybeacon.com	Conspiracy	192	truthfeed.com	Hate
43	redrocktribune.com	Fake News	93	loanpride.com	Rumor	143	investmentresearchdynamics.com	Conspiracy	193	darkmoon.me	Hate
44	uspoliticlive.com	Fake News	94	newswire-24.com	Rumor	144	truthbroadcastnetwork.com	Conspiracy	194	glaringhypocrisy.com	Hate
45	americannews.com	Fake News	95	collectivelyconscious.net	Junk	145	thephaser.com	Conspiracy	195	wsj.com	Reliable
46	universepolitics.com	Fake News	96	naturalnews.com	Science	146	jesus-is-savior.com	Conspiracy	196	nytimes.com	Reliable
47	dailyheadlines.net	Fake News	97	naturalnewsblogs.com	Science	147	rense.com	Conspiracy	197	theguardian.com	Reliable
48	religionmind.com	Fake News	98	icr.org	Science	148	themindunleashed.com	Conspiracy			
49	conservativefighters.com	Fake News	99	thetruthaboutcancer.com	Science	149	jihadwatch.org	Conspiracy			
50	viralactions.com	Fake News	100	ancient-code.com	Science	150	intellihub.com	Conspiracy			

**Table 2 Comparison with other datasets and papers.**

Dataset	Instances	Categories	Av. Words per Instance	Source
Kaggle Fake News	12,999	1	637	(Risidal, 2016)
Fake News Challenge	49,974	4	11	(Rubin et al., 2015)
LIAR	12,791	6 <sup>a</sup>	18	(Wang, 2017)
Univ. of Washington Fake News Data	60,841	4	530	(Rashkin et al., 2017)
The Fingerprints of Misinformation	92,112	7	461	

<sup>a</sup>On a scale from 0 to 5; 0 being completely false to 5 being completely accurate

In comparison to other datasets, the one used in this paper includes more sources and more articles than any other:

**Computational linguistics.** Extant research in the human cognition and behavioral sciences can be leveraged to identify misinformation online through quantitative measures. For example, the information manipulation theory (McCornack et al., 2014) or the four-factor theory (Zuckerman et al., 1981) propose that misinformation is expressed differently in terms of arousal, emotions or writing style. The main intuition is that misinformation creators have a different writing style seeking to maximize reading, sharing and, in general, maximizing virality. This is important because views and engagement in social networks are closely related to virality, and being repeatedly exposed to misinformation increases the likelihood of believing in false claims (Bessi et al., 2015; Mocanu et al., 2015).

Based on the information manipulation theory and the four-factor theory, I propose several parametrizations that can allow to statistically distinguish between factual news and a myriad of misinformation categories. To do so, I calculate a set of quantifiable characteristics that represent the content of a written text and allow us to differentiate it across categories. More specifically, following the definition proposed by (Zhou & Zafarani, 2020) the style-based categorization of content is formulated as a multinomial classification problem. In this type of problem, each text in a set of news articles  $N$  can be represented as a set of  $k$  features denoted by the feature vector  $f \in \mathbb{R}^k$ . Through Natural Language Processing, or computational linguistics, I can calculate this set of  $k$  features for  $N$  texts. In the following subsection I describe these features and their theoretical grounding.

**Measuring cognitive effort through grammatical features: readability.** Using sentence length or word syllables as a measure of text complexity has a long tradition in computational linguistics (Afroz et al., 2012; Fuller et al., 2009; Hauch et al., 2015; Monteiro et al., 2018). Simply put, the longer a sentence or word is, the more complex it is to read. Therefore, longer sentences and longer words require more cognitive effort to be effectively processed. The length of sentences and words are precisely the fundamental parameters of the Flesch-Kincaid index, the readability variable.

Flesch-Kincaid is used in different scientific fields like pediatrics (D'Alessandro et al., 2001), climate change (De Bruin & Granger Morgan, 2019), tourism (Liu & Park, 2015) or social media (Rajadesingan et al., 2015). This measure estimates the educational level that is needed to understand a given text. The Flesch-Kincaid readability score is calculated with the following formula (Kincaid et al., 1975):

$$\text{Flesch.Kincaid score(FK)} = 0.39 * \left( \frac{n_w}{n_{st}} \right) + 11.8 * \frac{n_{sy}}{n_w} - 15.59 \quad (1)$$

where  $n_w$  is the number of words,  $n_{st}$  is the number of sentences, and  $n_{sy}$  is the number of syllables. In this case,  $n_w$  and  $n_{st}$  act as a proxy for syntactic complexity and  $n_{sy}$  acts as a proxy for lexical difficulty. All of them are important components of readability (Just & Carpenter, 1980).

While the Flesch-Kincaid score measured the cognitive effort needed to process a text based on grammatical features (the number of words, sentences, and syllables), it does not account for another source of cognitive load: lexical features.

**Measuring cognitive effort through lexical features: perplexity.** Lexical diversity is defined as a measure of the number of different words used in a text (Beheshti et al., 2020). In general, more advanced and diverse language allows to encode more complex ideas (Ellis & Yuan, 2004), which generates a higher cognitive load (Swabey et al., 2016). One of the most obvious measures for lexical diversity is using the ratio of individual words to the total number of words (known as the type-token ratio or TTR). However, this measure is extremely influenced by the denominator (text length). Therefore, I calculate the uncertainty in predicting each word appearance in every text through perplexity (Griffiths & Steyvers, 2004), a measure that has been used for language identification (Gamallo et al., 2016), to discern between formal and informal tweets (González, 2015), to model children's early grammatical knowledge (Bannard et al., 2009), measuring the distance between languages (Gamallo et al., 2017) or to assess racial disparities in automated speech recognition (Koenecke et al., 2020).

For any given text, there is a probability  $p$  for each word to appear. Lower probabilities indicate more information while higher probabilities indicate less information. For example, the word "aerospace" (low probability) has more information than "the" or "and" (high probability). From here, I can calculate how "surprising" each word  $x$  is by using  $\log(p(x))$ . Therefore, words that are certain to appear have 0 surprise ( $p = 1$ ) while words that will never appear have infinite surprise ( $p = 0$ ). Entropy is the average amount of "surprise" per word in each text, therefore serves as a measure of uncertainty (higher lexical diversity) and it is calculated with the following formula:

$$H = - \sum_x p(x) \log_2 q(x) \quad (2)$$

where  $p(x)$  and  $q(x)$  are the probability of word  $x$  appearing in each text. The negative sign ensures that the result is always positive or zero. For example, a text containing the string "bla bla bla bla bla" has an entropy of 0 because  $p(bla) = 1$  (a certainty), while the string "this is an example of higher entropy" has an entropy of 2.807355 (higher uncertainty). Building upon entropy, perplexity measures the amount of "randomness" in a text:

$$\text{Perplexity}(M) = \tau^{- \sum_x p(x) \log_2 q(x)} = 2^H \quad (3)$$

where  $\tau$  in our case is 2, and the exponent is the cross-entropy. All else being equal, a smaller vocabulary generally yields lower perplexity, as it is easier to predict the next word in a sequence

(Koenecke et al., 2020). The interpretation is as follows: If perplexity equals 5, it means that the next word in a text can be predicted with an accuracy of 1-in-5 (or 20%, on average). Following the previous example, the string “bla bla bla bla bla” has a perplexity of 1 (no surprise because all words are the same and, therefore, predicted with a probability of 1), while the string “this is an example of higher entropy” has a perplexity of 7 (since there are 7 different words that appear 1 time each, yielding a probability of 1-in-7 to appear).

**Measuring emotions through polarity: sentiment analysis.** The usual way of measuring polarity in written texts is through sentiment analysis. For example, this technique has been used to analyze movie reviews (Bai, 2011), to improve ad relevance (Qiu et al., 2010), to quantify consumers’ ad sharing intentions (Kulkarni et al., 2020), to explore customer satisfaction (Ju et al., 2019) or to predict election outcomes (Tumasjan et al., 2010). Mining opinions in texts is done by seeking content that captures the effective meaning of sentences in terms of sentiment. In this case, I am interested in the determination of the emotional state (positive, negative, or neutral) that the text tries to convey towards the reader. To do so, I employ a dictionary to help in the achievement of this task. More specifically, I employ the AFINN lexicon developed by Finn Årup Nielsen (Nielsen, 2011) one of the most used lexicons for sentiment analysis (Akhtar et al., 2020; Chakraborty et al., 2020; Hee et al., 2018; Ragini et al., 2018). In this dictionary, 2,477 coded words have a score between minus five (negative) to plus five (positive). The algorithm matches words in the lexicon in each text and adds/subtracts points as it effectively finds positive and negative words in the dictionary that appear in the text. If a text has a neutral evocation to emotions will have a value around 0, if a text is evocating positive emotions will have a value higher than 0 and if a text is evocation a negative emotion will have a value below 0. For example, in my sample, one of the highest values in negative emotions (emotion = -32) is the following news from the fake news category reporting about a shooting against two police officers in France: “(...) Her candidacy [referring to Marine Le Pen] has been an uprising against the globalist-orchestrated Islamist invasion of the EU and the associated loss of sovereignty. The EU is responsible for the flood of terrorist and Islamists into France (...)”. In contrast, the following junk science article has one of the highest positive values (emotion = +28): “A new bionic eye lenses currently in development would give humans 3x vision, at any age. (...) Even better is the fact that people who get the lens surgically inserted will never develop cataracts”.

**Measuring emotions through social identity: morality.** To measure morality, I will use a previously validated dictionary (Graham et al., 2009). This dictionary has been used to measure polarizing topics like gun control, same-sex marriage or climate change (Brady et al., 2017), to study propaganda (Barrón-Cedeño et al., 2019) or to measure responses to terrorism (Sagi & Dehghani, 2014) and social distance (Dehghani et al., 2016). Like in sentiment analysis, morality is measured by counting the frequency of moral words in each text. The dictionary contains 411 words like “abomination”, “demon”, “honor”, “infidelity”, “patriot” or “wicked”. In contrast to previous measures, the technique employed to quantify morality is highly sensitive to text length (with longer texts having higher probabilities of containing “moral” words), therefore, I calculate the morality measure as moral words per 500 words in each text. In addition, since negative news spreads farther (Hansen et al., 2011; Vosoughi et al., 2018), I add an interaction term between morality and negativity by multiplying the morality per 500 words and the

negativity per 500 words for each text, this being the main measurement for morality:

$$Morality_i(M) = \frac{mor_i * neg_i}{n_{w,i} \quad n_{w,i}} \tag{4}$$

where  $mor_i$  is the overall number of moral words in text  $i$ ,  $neg_i$  is the absolute number of negative words in text  $i$  and  $n_{w,i}$  is the total number of words in text  $i$ .

**Similarities between misinformation and factual news: distance and clustering.** For the distance metric I use the Euclidean distance that can be formulated as follows:

$$d_{AB} = \sqrt{\sum_{i=1}^n (e_{Ai} - e_{Bi})^2} \tag{5}$$

Regarding the method to merge the sub-clusters in the dendrogram, I employed the unweighted pair group method with arithmetic mean. Here, the algorithm considers clusters  $A$  and  $B$  and the formula calculates the average of the distances taken over all pairs of individual elements  $a \in A$  and  $b \in B$ . More formally:

$$d_{AB} = \frac{\sum_{a \in A, b \in B} d(a, b)}{|A| * |B|} \tag{6}$$

To add robustness to the results, I also used the k-means algorithm, which calculates the total within-cluster variation as the sum of squared Euclidean distances (Hartiga & Wong, 1979):

$$W(C_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2 \tag{7}$$

where  $x_i$  is a data point belonging to the cluster  $C_k$ , and  $\mu_k$  is the mean value of the points assigned to the cluster  $C_k$ . With this, the total within-cluster variation is defined as:

$$tot.withiness = \sum_{k=1}^k W(C_k) = \sum_{k=1}^k \sum_{x_i \in C_k} (x_i - \mu_k)^2 \tag{8}$$

To select the number of clusters, I use the elbow method with the aim to minimize the intra-cluster variation:

$$minimize \left( \sum_{k=1}^k W(C_k) \right) \tag{9}$$

**Differentiating misinformation from factual news: multinomial logistic regression.** The main objective of this paper is not just to report descriptive differences between reliable news and misinformation sources, but to look for systemic variance among their structural features measured through the four variables. Therefore, it is not enough to report the averages and confidence intervals of each variable for each category, but also analyzing differences and similarities in the light of all variables altogether. This is why I will employ a multinomial logistic regression, a technique suitable for mutually exclusive variables with multiple discrete outcomes.

I employ a multinomial logistic regression model with  $K$  classes using a neural network with  $K$  outputs and the negative conditional log-likelihood (Venables & Ripley, 2002). This logistic model is generalizable to categorical variables with more than two levels namely  $\{1, \dots, J\} \{1, \dots, J\}$ . Given the predictors  $X_1, \dots, X_p, X_1, \dots, X_p$ . In this multinomial logistic regression model, the probability of each level  $j$  of  $Y$  is calculated with the

following formula (García-Portugés, 2021):

$$p_j(x) := \frac{\mathbb{P}\left[Y = j | X_1 = x_1, \dots, X_p = x_p\right] e^{\beta_{0j} + \beta_{1j}x_1 + \dots + \beta_{pj}x_p}}{1 + \sum_{l=1}^{J-1} e^{\beta_{0l} + \beta_{1l}x_1 + \dots + \beta_{pl}x_p}} \tag{10}$$

for  $j = 1, \dots, J - 1$   $l = 1, \dots, J - 1$  and (for the reference level  $J = \text{factual news}$ ):

$$p_j(x) := \frac{\mathbb{P}\left[Y = J | X_1 = x_1, \dots, X_p = x_p\right] 1}{1 + \sum_{l=1}^{J-1} e^{\beta_{0l} + \beta_{1l}x_1 + \dots + \beta_{pl}x_p}} \tag{11}$$

As a generalization of a logistic model, it can be interpreted in similar terms if we take the quotient between (A.1) and (A.2):

$$\frac{p_j(X)}{p_l(X)} = e^{\beta_{0j} + \beta_{1j}x_1 + \dots + \beta_{pj}x_p} \tag{12}$$

for  $j = 1, \dots, J - 1$   $l = 1, \dots, J - 1$ . If we apply a logarithm to both sides, we obtain:

$$\log\left(\frac{p_j(X)}{p_l(X)}\right) = \beta_{0j} + \beta_{1j}x_1 + \dots + \beta_{pj}x_p \tag{13}$$

Therefore, multinomial logistic regression is a set of  $J - 1$  independent logistic regressions for the probability of  $Y = j$  versus the probability of the reference  $Y = J$ . In this case, I used factual news as the level of the outcome since I am interested how misinformation differs from this baseline using the following formula:

$$\log\left(\frac{P(\text{cat} = j = \text{fake})}{P(\text{cat} = J = \text{reliable})}\right) = \beta_{0j} + \beta_{1j}(r) + \beta_{1j}(p) + \beta_{1j}(s) + \beta_{1j}(m) \tag{14}$$

where  $s = \text{sentiment}$ ,  $m = \text{morality}$ ,  $r = \text{readability}$  and  $p = \text{perplexity}$ . This method will allow us, beyond the fingerprints of misinformation described before, to quantify the differences between factual news and non-factual content.

**Results**

To investigate the differences between factual news and misinformation, I analyze 92,112 news articles classified into seven categories: clickbait ( $n = 12,955$ ), conspiracy theories ( $n = 15,493$ ), fake news ( $n = 16,158$ ), hate content ( $n = 15,353$ ), junk science ( $n = 16,252$ ), rumors ( $n = 14,158$ ), and factual information ( $n = 1743$ ) (see Methods for a detailed description of the database). For each article, I calculated measures of cognitive effort (readability and perplexity) and evocation to emotions (sentiment analysis and appeal to morality). Figure 1 shows the average values and confidence intervals for each measure and each category.

With regard to cognitive effort, I obtained the following results. The readability scores are high for junk science (FK = 13.94, [CI = 13.90, 13.99]). Hate content (FK = 12.97, [CI = 12.93, 13.03]) and factual news (FK = 12.96, [CI = 12.83, 13.09]) are very similar. A third group of news category contains clickbait (FK = 12.40, [CI = 12.35, 12.45]), conspiracy theories (FK = 12.39, [CI = 12.34, 12.43]) and rumors (FK = 12.37, [CI = 12.33, 12.42]). Finally, fake news is the category with a lower cognitive effort needed to process the content in all the categories as measured by the readability score (FK = 11.25, [CI = 11.21, 11.29]). Next, I examined perplexity with the following results. Rumors obtained the lowest value ( $p = 140.56$ , [CI = 140.12, 141.00]), followed by fake news ( $p = 142.56$ , [CI = 142.13, 142.99]), hate content ( $p = 143.64$ , [CI = 143.18, 144.10]) and conspiracy theories ( $p = 143.73$ , [CI = 143.28, 144.17]). Then, clickbait ( $p = 155.07$ , [CI = 154.57, 155.57]) and

junk science ( $p = 163.91$ , [CI = 163.53, 164.30]) are the categories right below factual content ( $p = 174.23$ , [CI = 172.99, 175.47]).

The sentiment analysis variable highlights how factual news are, in essence, neutral (sentiment = -0.19, [CI = -0.84, 0.45]), in concordance with its journalistic norm of objectivity (Neuman et al., 1992). On the side of more positive language, I find rumors (sentiment = 0.86, [CI = 0.65, 1.08]) and junk science (sentiment = 2.52, [CI = 2.31, 2.72]). Looking at categories with a negative prominence in their content, I find that hate has the highest negative value (sentiment = -6.01, [CI = -6.21, -5.82]), followed by fake news (sentiment = -3.51, [CI = -3.70, -3.32]), conspiracy theories (sentiment = -3.41, [CI = -3.59, -3.22]) and clickbait (sentiment = -3.09, [CI = -3.31, -2.87]). Regarding social identity and the appealing to morality, factual news has the lowest value of all categories (morality = 3.12, [CI = 3.01, 3.22]), again, in concordance with objectivity approaches in reliable news. Somehow in-between, I find categories that employ moral language to a greater extent but without high values. That is the case of junk science (morality = 3.87, [CI = 3.83, 3.91]) and rumors (morality = 3.97, [CI = 3.93, 4.02]). As for the categories with higher usage of moral language, there are clickbait news (morality = 4.38, [CI = 4.34, 4.43]), hate speech (morality = 4.41, [CI = 4.36, 4.46]), conspiracy theories (morality = 4.44, [CI = 4.39, 4.48]) and, finally, fake news (morality = 4.66, [CI = 4.61, 4.70]).

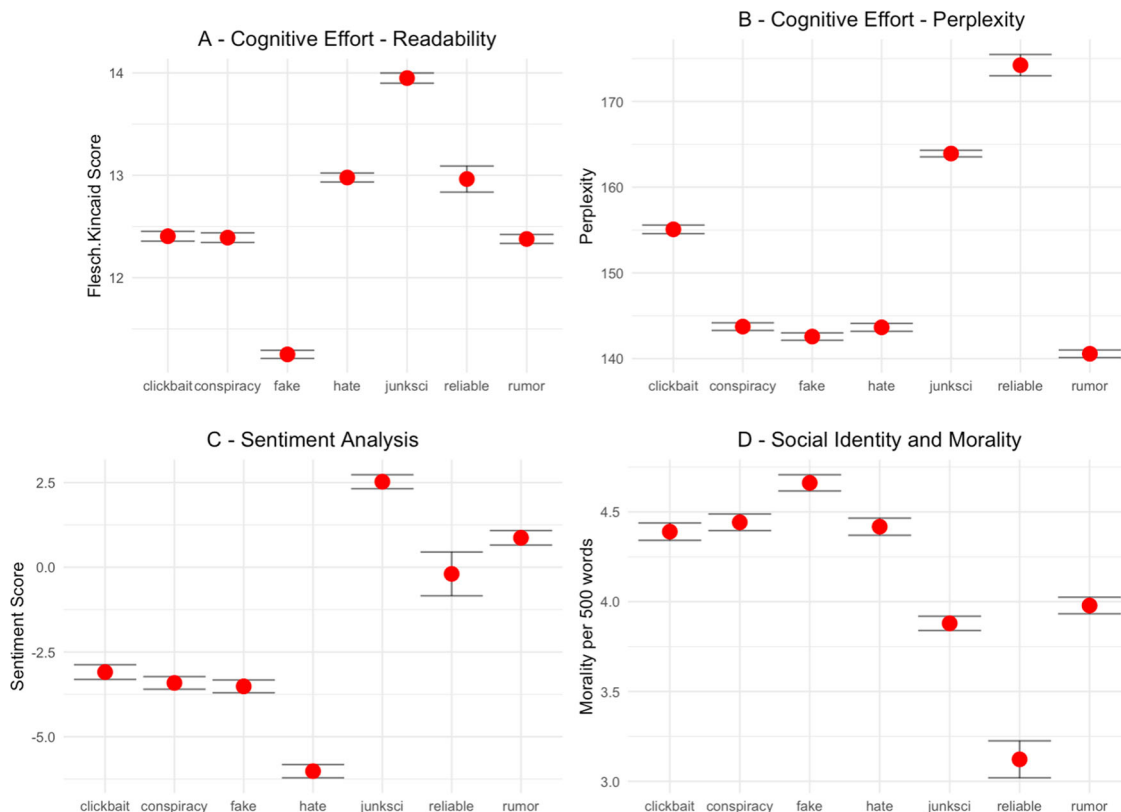
**Quantitative similarities among factual news and misinformation categories.**

After calculating the factual and misinformation categories profiles, I calculate similarities among them through clustering analysis. This is important because results will indicate how close each misinformation category is to reliable news, allowing us to refine the previously presented results. First, I calculated the Euclidean distance (sum of squared distances and taking the square root of the resulting value), forcing values that are very different to add a higher contribution to the distance among observations. After calculating the distance matrix, I employed a hierarchical clustering technique using the unweighted pair group method with arithmetic mean. The results are the following:

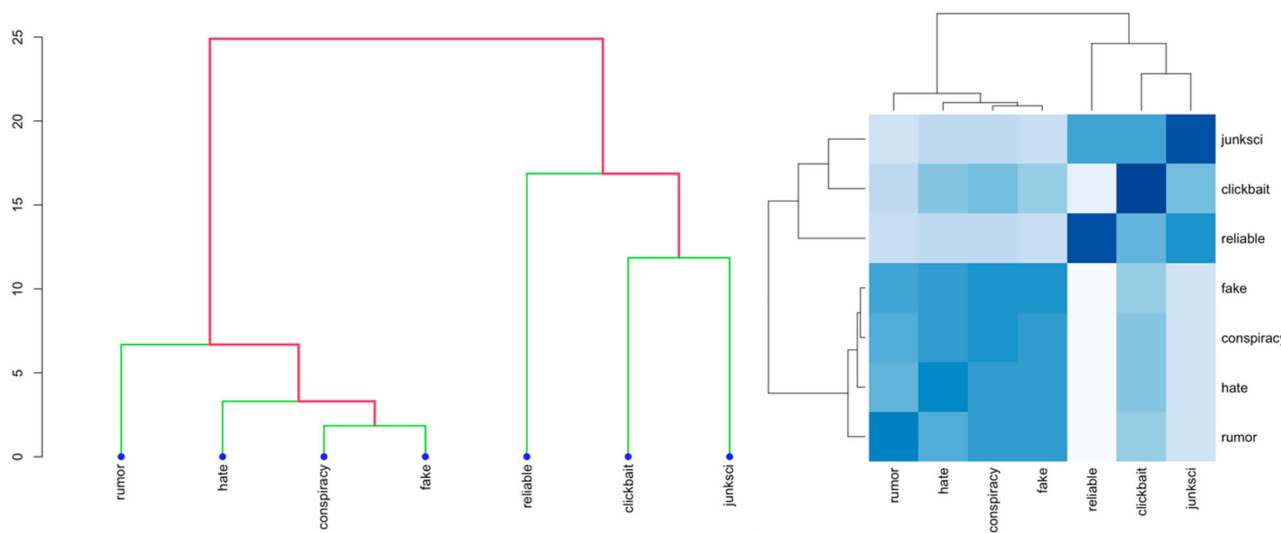
In Fig. 2A, B one can observe that the Euclidean distance between factual news and rumors is 37.68, with fake news is 35.69, with hate content is 34.84, with conspiracy theories is 34.33, with clickbait is 21.71 and with junk science is 12.00. Looking at the clustering resulting from these distances, one can see that there are two big clusters ( $height = 24.90$ ): rumors, hate speech, conspiracy theories and fake news on one side and factual content, clickbait, and junk science on the other. This result indicates that the misinformation categories that are more similar to reliable news are click bait and junk science. However, looking at the resulting clusters for a lower height, factual content is the first category to be separated from the others ( $height 16.86$ ). In other words, at height = 16.86, the biggest difference is between factual content and misinformation sources. From there, the next separation is between clickbait and rumors ( $height = 11.85$ ). On the other hand, the separations in the first group appear at lower heights (meaning, more similarities among these categories). For example, rumors split from hate speech, conspiracy theories and fake news at height = 6.68. Next, hate speech separates from conspiracy theories and fake news at height = 3.30. Finally, the most similar categories are conspiracy theories and fake news ( $height = 1.84$ ).

Using the total within-clusters sum of squares in the sample, one can observe that the optimal number of clusters are two. However, I also report all the other clustering possibilities as a robustness check.





**Fig. 1 The fingerprints of misinformation.** **A** shows the cognitive effort needed to process a text using the Flesch-Kincaid readability score; **B** the cognitive effort as measured by perplexity; **C** plots the sentiment (positive, neutral, negative) of each category; **D** shows the appeal to morality in each category.

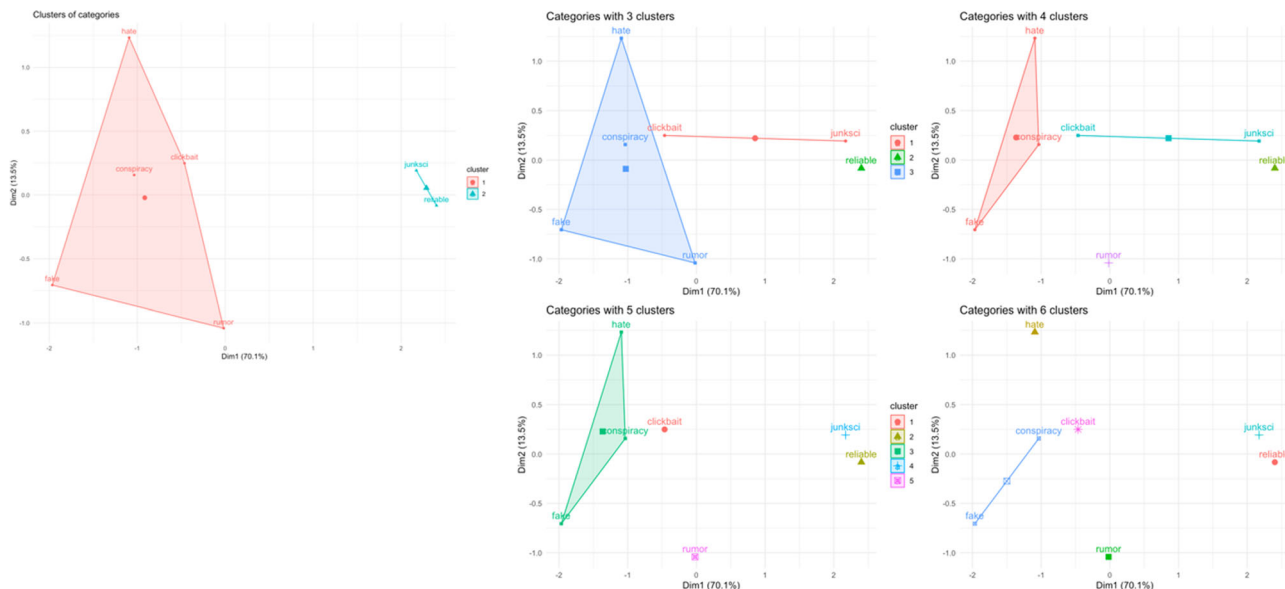


**Fig. 2 Cluster similarities.** Hierarchical clustering of categories. **A** red lines indicate stronger associations. **B** darker colors indicate stronger associations.

In Fig. 3A, junk science and factual content are the most similar categories, while all the rest pertain to a single big cluster. However, if one opt for explore the results increasing the number of clusters (Fig. 3B), they follow the same behavior as the hierarchical clustering. After three clusters, reliable news is the first category to be isolated from all the others, revealing its distinctive nature in terms of linguistic characteristics.

**Quantitative differences among factual news and misinformation categories.** In Table 3 and Fig. 4 one can see the

results of the multinomial logistic regression model with each column comparing the corresponding misinformation category to the baseline (factual news). This method allows us to use reliable news as a “role model” of information and see how misinformation categories diverge from this baseline. The results show that a one-unit increase in the readability score is associated with a decrease in the log odds of categorizing content in the clickbait ( $\beta_{\text{readability, clickbait}} = -0.06, p < 0.001$ ) conspiracy theories ( $\beta_{\text{readability, conspiracy}} = -0.05, p < 0.001$ ), fake news ( $\beta_{\text{readability, fake news}} = -0.21, p < 0.001$ ), and rumor ( $\beta_{\text{readability, rumor}} = -0.04, p < 0.001$ ) categories. In other words, the



**Fig. 3 Clustering analysis.** **A** Clustering results with two clusters. **B** Clustering results with more than two clusters.

**Table 3 Quantitative differences among factual news and misinformation categories.**

	Clickbait	Conspiracy	Fake News	Hate Speech	Junk Science	Rumor
(Intercept)	6335*** (0.057)	8.575*** (0.052)	10.475*** (0.053)	7.859*** (0.052)	2.509*** (0.061)	9.176*** (0.054)
Readability	-0.059*** (0.008)	-0.053*** (0.008)	-0.213*** (0.008)	0.014+ (0.008)	0.127*** (0.008)	-0.044*** (0.008)
Perplexity	-0.026*** (0.001)	-0.040*** (0.001)	-0.041*** (0.001)	-0.041*** (0.001)	-0.015*** (0.001)	-0.045*** (0.001)
Sentiment	-0.013*** (0.002)	-0.016*** (0.002)	-0.018*** (0.002)	-0.033*** (0.002)	0.024*** (0.002)	0.010*** (0.002)
Morality	0.182*** (0.012)	0.179*** (0.012)	0.197 (0.012)	0.159*** (0.012)	0.160*** (0.012)	0.144*** (0.012)
Num. obs.	92112					
AIC	318785.7					
edf	30.000					

†p < 0.1, \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001.

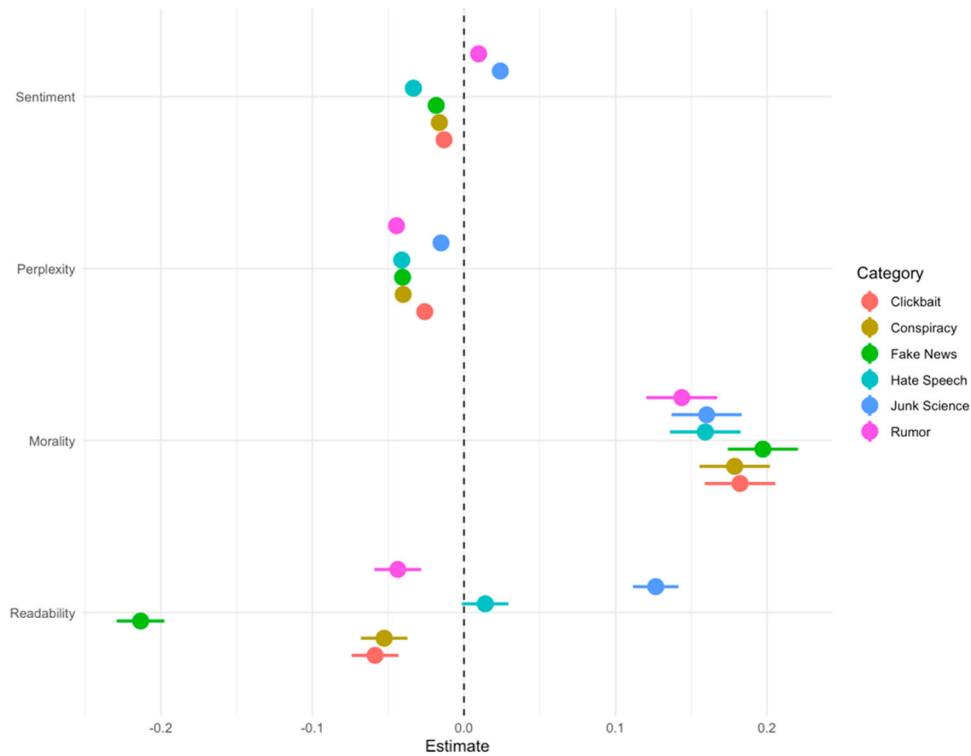
easier to read a text is, the more likely it pertains to the clickbait, conspiracy theories, fake news, or rumor categories. On the other side, a one-unit increase in the Flesch-Kincaid score is associated with an increase in the log odds of hate speech ( $\beta_{\text{readability, hate}} = 0.01, p < 0.1$ ) and junk science categories ( $\beta_{\text{readability, junk science}} = 0.04, p < 0.001$ ).

Regarding perplexity, the log odds of a content being misinformation ( $\beta_{\text{perplexity, clickbait}} = -0.03, p < 0.001$ ;  $\beta_{\text{perplexity, conspiracy}} = -0.04, p < 0.001$ ;  $\beta_{\text{perplexity, fake news}} = -0.04, p < 0.001$ ;  $\beta_{\text{perplexity, hate}} = -0.04, p < 0.001$ ;  $\beta_{\text{perplexity, junk science}} = -0.01, p < 0.001$ ;  $\beta_{\text{perplexity, conspiracy}} = -0.04, p < 0.001$ ) decreases as the level of perplexity of the text increases (as a reminder, an increase in perplexity means lower predictability). In other words, misinformation categories have lower lexical diversity.

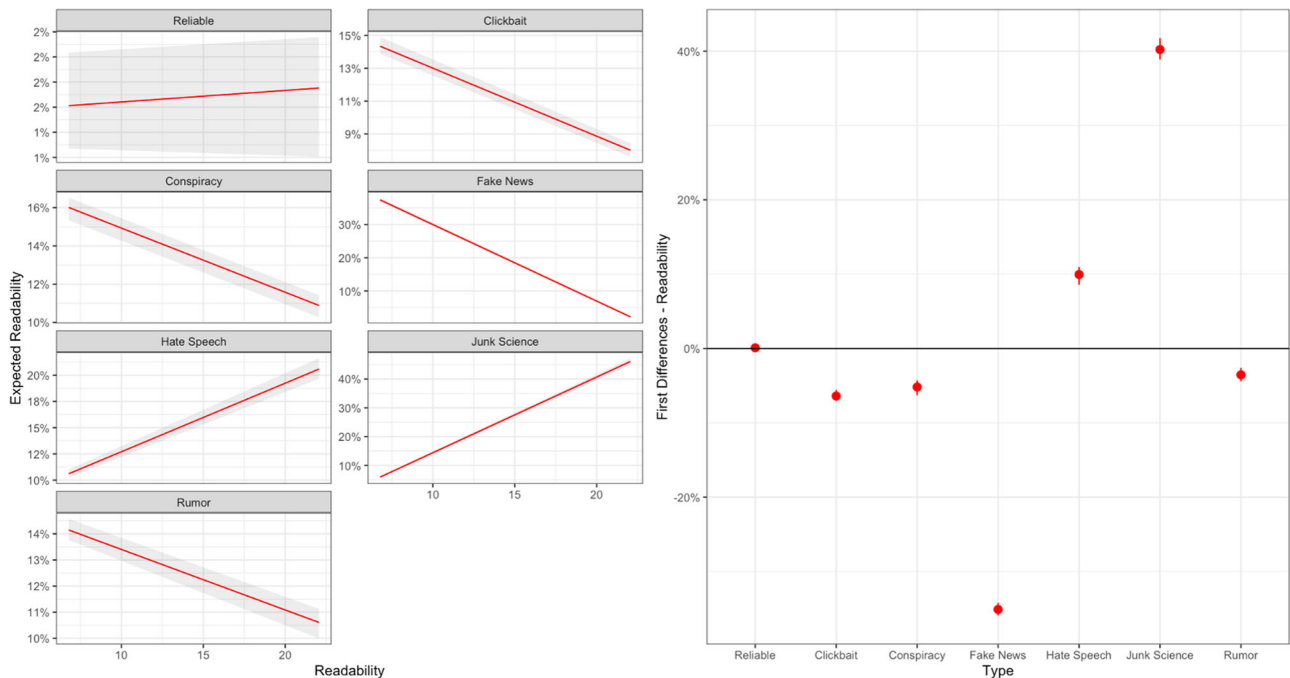
As for the sentiment of the content, a one-unit increase in negative sentiment is associated with an increase in the log odds of a content being clickbait ( $\beta_{\text{sentiment, clickbait}} = -0.01, p < 0.001$ ), conspiracy theory ( $\beta_{\text{sentiment, conspiracy}} = -0.02, p < 0.001$ ), fake news ( $\beta_{\text{sentiment, fake news}} = -0.02, p < 0.001$ ) or hate speech ( $\beta_{\text{sentiment, hate}} = -0.03, p < 0.001$ ), indicating that these categories

tend to employ a highly negative and sentimental language; while a one-unit increase in the positive sentiment score is associated with an increase in the log odds of a content being junk science ( $\beta_{\text{sentiment, junk science}} = 0.02, p < 0.001$ ) or a rumor ( $\beta_{\text{sentiment, rumor}} = 0.01, p < 0.001$ ). Therefore, misinformation tends to rely on emotional language. However, the polarity of these emotions varies across misinformation categories. Both junk science and rumors tend to be significantly more positive than reliable news.

Finally, an increase by one-unit in the morality appealing in a given text is associated with an increase in the log odds of this content being misinformation ( $\beta_{\text{morality, clickbait}} = 0.18, p < 0.001$ ;  $\beta_{\text{morality, conspiracy}} = 0.18, p < 0.001$ ;  $\beta_{\text{morality, fake news}} = 0.20, p < 0.001$ ;  $\beta_{\text{morality, hate}} = 0.16, p < 0.001$ ;  $\beta_{\text{morality, junk science}} = 0.16, p < 0.001$ ;  $\beta_{\text{morality, conspiracy}} = 0.14, p < 0.001$ ). In addition to lexical diversity, the usage of moral language appears to be one of the main determinants of misinformation communication strategies. This finding is important because existing research tends to overemphasize the role of sentiment while neglecting the prominent role of appeal to morality.



**Fig. 4 Multinomial logistic regression results.** Factual news are considered the level of the model output (i.e., the reference).

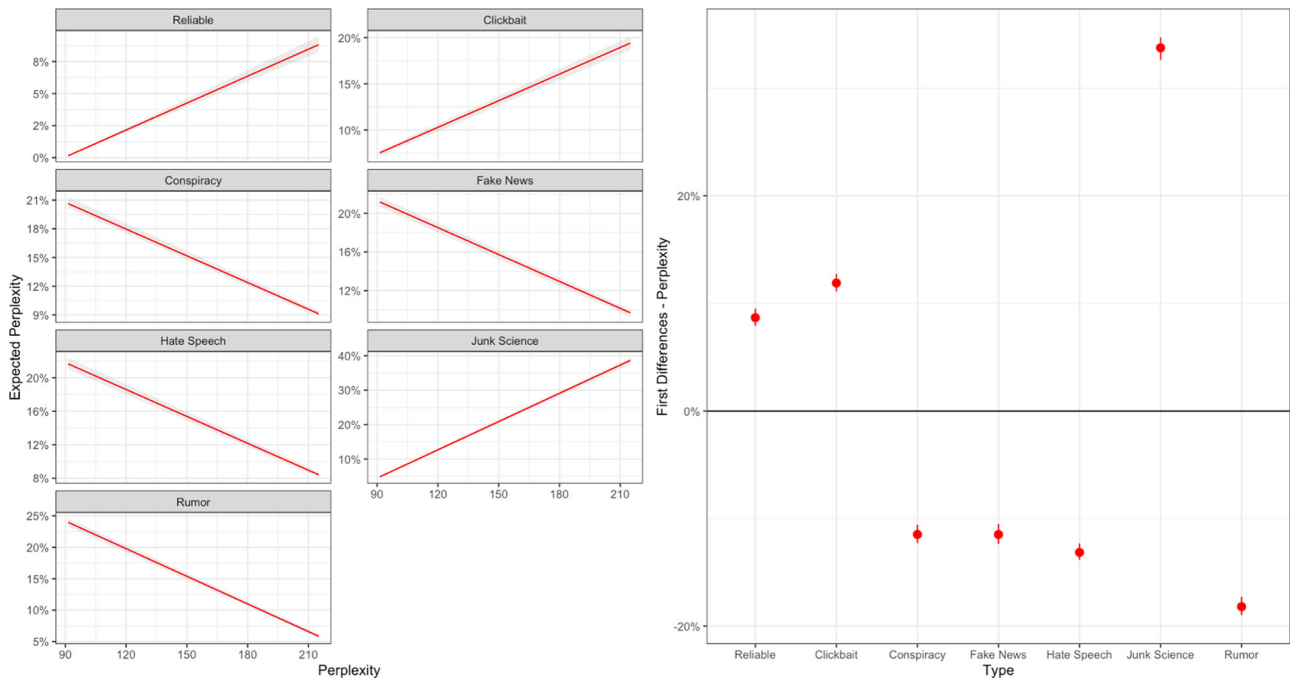


**Fig. 5 First differences of the multinomial logit model – Readability.**

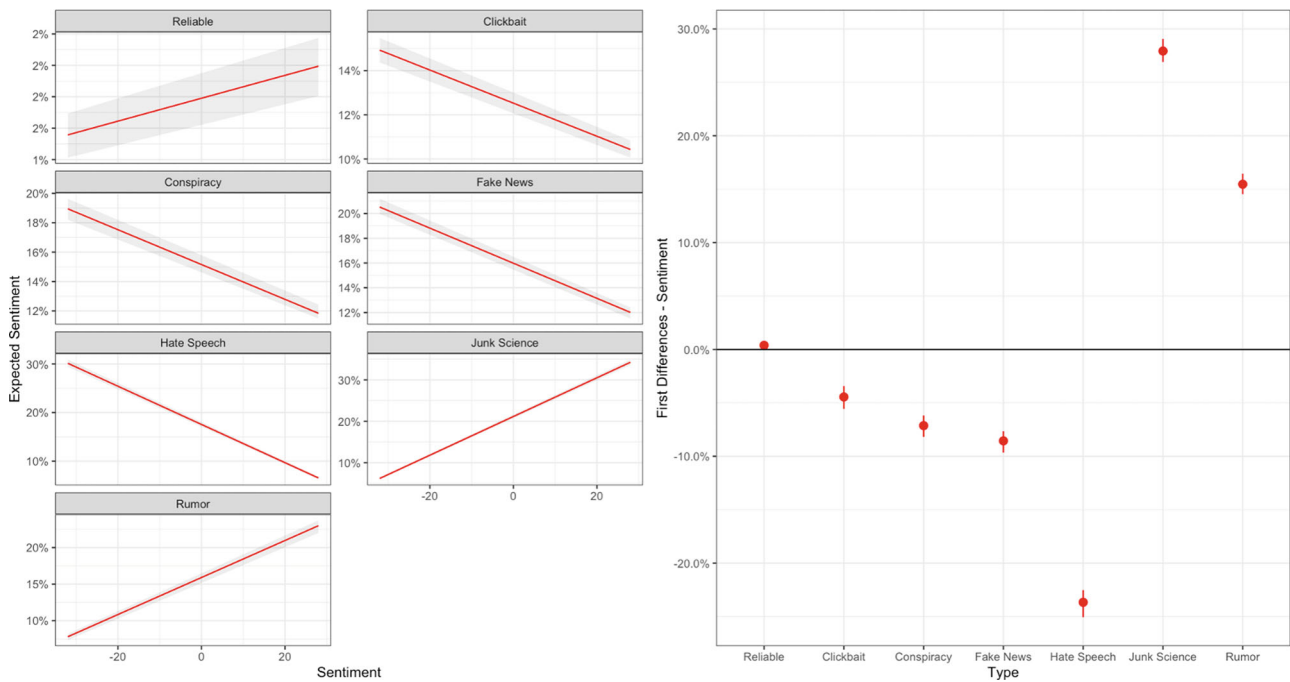
The results of the multinomial logistic regression were used to get insights into the probabilities of categorizing content using a baseline category (factual news). Next, I complement the results with predicted probabilities based on simulations for each category in four scenarios corresponding to the four variables using the previous multinomial logistic regression model. In other words, one can predict the probabilities for each category in the scenario (variable) with the following ranges: Readability = [6.78, 22.05],

Perplexity = [91.37, 215.01], Sentiment = [-32, 28], Mortality = [0.13, 13.75]. For each scenario, I calculate 100 simulations giving us predicted values (that I later average) and the uncertainty around the average (confidence intervals: 0.025, 0.975).

Regarding the first differences, I use the expected values. The difference between predicted and expected values is subtle but important. Even though both result in almost identical averages ( $r = 0.999$ ,  $p < 0.001$ ), predicted values have a larger variance



**Fig. 6** First differences of the multinomial logit model – Perplexity.



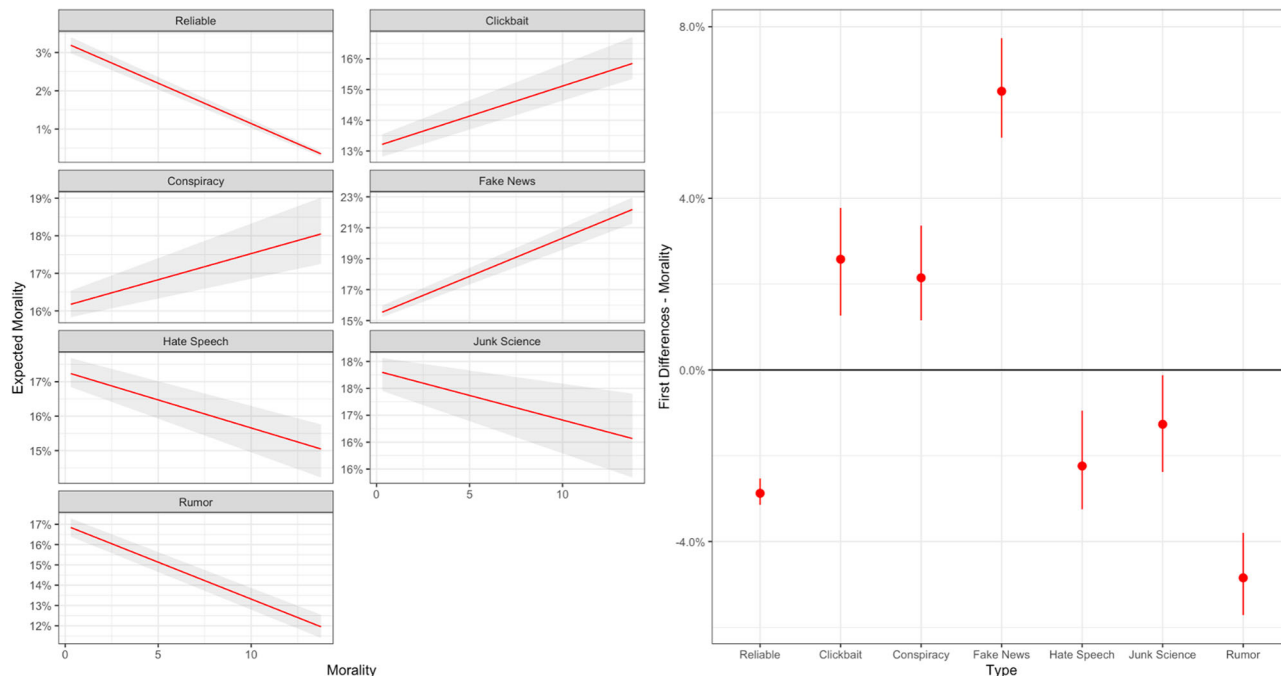
**Fig. 7** First differences of the multinomial logit model – Sentiment.

because they content both fundamental and estimation uncertainty (King, Tomz, & Wittenberg, 2000). Therefore, to calculate the expected value I apply the same procedure for the predicted value and average over the fundamental uncertainty of the  $m$  simulations (in this case,  $m = 100$ ). Specifically, the procedure is to simulate each variable setting all the other variables at their means and the variable I am interested in at its starting point (lower range). Then, I change the value of the variable to its ending point (high range), keeping all the other variables at their means and repeat the simulation. I repeat this process 100 times and average the results for the starting and

ending points. The result is a full probability distribution that I use to compute the average expected value and the confidence intervals. From there, I calculate first differences, which are the difference between the two expected, rather than predicted, values (King et al., 2000).

These results show the predicted probabilities for all choices of the multinomial logit model I employed. For each variable I performed 100 simulations with confidence intervals settled at 0.025 and 0.975. The results are the following (Figs. 5–8):

For lower levels of complexity (readability = 6.78), the probabilities for a given text being classified in each category is:



**Fig. 8** First differences of the multinomial logit model – Morality.

factual ( $p = 0.015$ ,  $CI = [0.013, 0.017]$ ), clickbait ( $p = 0.145$ ,  $CI = [0.140, 0.149]$ ), conspiracy theories ( $p = 0.160$ ,  $CI = [0.156, 0.166]$ ), fake news ( $p = 0.373$ ,  $CI = [0.366, 0.380]$ ), hate speech ( $p = 0.106$ ,  $CI = [0.102, 0.109]$ ), junk science ( $p = 0.060$ ,  $CI = [0.058, 0.061]$ ) and rumor ( $p = 0.141$ ,  $CI = [0.136, 0.147]$ ). In contrast, for complex texts (readability = 22), the probabilities per category are: factual ( $p = 0.015$ ,  $CI = [0.013, 0.018]$ ), clickbait ( $p = 0.079$ ,  $CI = [0.074, 0.084]$ ), conspiracy theories ( $p = 0.109$ ,  $CI = [0.104, 0.115]$ ), fake news ( $p = 0.022$ ,  $CI = [0.020, 0.023]$ ), hate speech ( $p = 0.106$ ,  $CI = [0.199, 0.214]$ ), junk science ( $p = 0.461$ ,  $CI = [0.452, 0.471]$ ) and rumor ( $p = 0.105$ ,  $CI = [0.099, 0.112]$ ). Regarding first differences, one can see that increasing the readability score from 6.78 to 22 has no effect on the probabilities of content being classified as factual news ( $p = 0.000$ ,  $CI = [-0.003, 0.003]$ ), decrease by 6.44% the probabilities of being clickbait ( $CI = [-0.074, -0.056]$ ), a decrease of 5.16% for conspiracy theories ( $CI = [-0.061, -0.042]$ ), or a decrease of 3.47% in rumors ( $CI = [-0.044, -0.026]$ ); remarkably, increasing the readability translates into a decrease of 35% of being classified as fake news ( $CI = [-0.361, -0.344]$ ). On the other side, an increase in readability shows 9.96% more probability of being classified as hate speech ( $CI = [0.087, 0.111]$ ) and an increase of 40.19% of being classified as junk science ( $CI = [0.390, 0.414]$ ). These results show that the junk science and fake news categories are both the most sensible to changes in the readability score in opposite directions: increasing the readability score significantly increases the probability of being categorized as junk science while decreasing it significantly increases the probability of a given content being categorized as fake news.

Next, for the cognitive effort measured through perplexity, one can see that lower levels of perplexity (perplexity = 91), the probabilities for a given text being classified in each category is: factual ( $p = 0.001$ ,  $CI = [0.001, 0.002]$ ), clickbait ( $p = 0.076$ ,  $CI = [0.073, 0.079]$ ), conspiracy theories ( $p = 0.206$ ,  $CI = [0.201, 0.212]$ ), fake news ( $p = 0.213$ ,  $CI = [0.208, 0.218]$ ), hate speech ( $p = 0.216$ ,  $CI = [0.211, 0.222]$ ), junk science ( $p = 0.049$ ,  $CI = [0.046, 0.050]$ ) and rumor ( $p = 0.239$ ,  $CI = [0.235, 0.246]$ ). In contrast, for content with high perplexity (perplexity = 215), the

probabilities per category are: factual ( $p = 0.089$ ,  $CI = [0.083, 0.096]$ ), clickbait ( $p = 0.194$ ,  $CI = [0.186, 0.200]$ ), conspiracy theories ( $p = 0.091$ ,  $CI = [0.088, 0.094]$ ), fake news ( $p = 0.098$ ,  $CI = [0.094, 0.101]$ ), hate speech ( $p = 0.084$ ,  $CI = [0.081, 0.088]$ ), junk science ( $p = 0.385$ ,  $CI = [0.376, 0.393]$ ) and rumor ( $p = 0.058$ ,  $CI = [0.056, 0.061]$ ). For perplexity, the first differences show that an increase from 91.37 to 215.01 translates into an 8.72% increase in the probabilities of content being classified as factual news ( $CI = [0.080, 0.095]$ ), an increase in 11.85% for the probability of being clickbait ( $CI = [0.109, 0.127]$ ), and an increase in 33.75% of being classified as junk science ( $CI = [0.327, 0.350]$ ). In contrast, it also translates into a 11.44% decrease in being classified as a conspiracy theory ( $CI = [-0.122, -0.107]$ ), a 11.45% decrease of being fake news ( $CI = [-0.123, -0.105]$ ), a 13.15 decrease in being hate speech ( $CI = [-0.140, -0.123]$ ) and a 18.27% decrease of being a rumor ( $CI = [-0.193, -0.175]$ ). These results indicate that increasing the perplexity of content increases the probabilities of a content being classified as junk science, clickbait or factual news while decreasing perplexity augments the probability of a piece of text being classified as conspiracy theory, fake news, hate speech or rumors.

As for sentiment analysis, highly negative values (sentiment = -32) give the following probabilities for each category: factual ( $p = 0.015$ ,  $CI = [0.013, 0.016]$ ), clickbait ( $p = 0.149$ ,  $CI = [0.143, 0.155]$ ), conspiracy theories ( $p = 0.189$ ,  $CI = [0.182, 0.195]$ ), fake news ( $p = 0.205$ ,  $CI = [0.201, 0.212]$ ), hate speech ( $p = 0.301$ ,  $CI = [0.292, 0.308]$ ), junk science ( $p = 0.062$ ,  $CI = [0.060, 0.065]$ ) and rumor ( $p = 0.077$ ,  $CI = [0.074, 0.080]$ ). In contrast, for highly positive sentiments (sentiment 28), the probabilities per category are: factual ( $p = 0.019$ ,  $CI = [0.017, 0.020]$ ), clickbait ( $p = 0.104$ ,  $CI = [0.100, 0.108]$ ), conspiracy theories ( $p = 0.118$ ,  $CI = [0.113, 0.123]$ ), fake news ( $p = 0.120$ ,  $CI = [0.115, 0.125]$ ), hate speech ( $p = 0.066$ ,  $CI = [0.063, 0.068]$ ), junk science ( $p = 0.342$ ,  $CI = [0.333, 0.348]$ ) and rumor ( $p = 0.231$ ,  $CI = [0.225, 0.239]$ ). In the case of sentiment, first differences show that increasing the sentiment score from -32 to 28 generates a negligible effect on content being classified as factual news ( $p = 0.000$ ,  $CI = [0.000, 0.007]$ ), and an increase in the probabilities of classifying content into the junk science ( $p = 0.279$ ,  $CI = [0.270, 0.290]$ ) and rumor

categories ( $p = 0.154$ ,  $CI = [0.141, 0.165]$ ). On the other side, an increase towards negative sentiments translates into an increase in probabilities of being categorized as clickbait ( $p = -0.044$ ,  $CI = [-0.053, -0.036]$ ), conspiracy theories ( $p = -0.071$ ,  $CI = [-0.082, -0.059]$ ), fake news ( $p = -0.085$ ,  $CI = [-0.096, -0.075]$ ), and hate speech ( $p = -0.236$ ,  $CI = [-0.247, -0.223]$ ). These results indicate that highly positive sentiments increase the probability of a content being classified as junk science or rumors, while an increase in the negative sentiment leads to an increased probability of being classified as hate speech, fake news, conspiracy theories or clickbait.

Finally, looking at low levels of appeal to morality (morality = 0.3), the probabilities for a given text being classified in each category is: factual ( $p = 0.032$ ,  $CI = [0.013, 0.016]$ ), clickbait ( $p = 0.132$ ,  $CI = [0.143, 0.155]$ ), conspiracy theories ( $p = 0.161$ ,  $CI = [0.182, 0.195]$ ), fake news ( $p = 0.156$ ,  $CI = [0.201, 0.212]$ ), hate speech ( $p = 0.172$ ,  $CI = [0.292, 0.308]$ ), junk science ( $p = 0.179$ ,  $CI = [0.060, 0.065]$ ) and rumor ( $p = 0.168$ ,  $CI = [0.074, 0.080]$ ). In contrast, for content with high appeal to morality (morality = 13.75), the probabilities per category are: factual ( $p = 0.003$ ,  $CI = [0.003, 0.004]$ ), clickbait ( $p = 0.158$ ,  $CI = [0.150, 0.166]$ ), conspiracy theories ( $p = 0.182$ ,  $CI = [0.174, 0.191]$ ), fake news ( $p = 0.221$ ,  $CI = [0.212, 0.230]$ ), hate speech ( $p = 0.151$ ,  $CI = [0.144, 0.158]$ ), junk science ( $p = 0.165$ ,  $CI = [0.158, 0.171]$ ) and rumor ( $p = 0.119$ ,  $CI = [0.112, 0.125]$ ). Finally, first differences for appealing to morality show that increasing the morality score from 0.3 to 13.75 generates a decrease on content being classified as factual news ( $p = -0.029$ ,  $CI = [-0.032, -0.025]$ ), a decrease in being classified as junk science ( $p = -0.013$ ,  $CI = [-0.032, -0.025]$ ), a decrease in hate speech ( $p = -0.022$ ,  $CI = [-0.032, -0.009]$ ), and a decrease in being a rumor ( $p = -0.050$ ,  $CI = [-0.058, -0.040]$ ). Contrarily, the same change in the scenario is associated with an increase being classified as conspiracy theory ( $p = 0.021$ ,  $CI = [0.009, 0.033]$ ), clickbait ( $p = 0.026$ ,  $CI = [0.012, 0.040]$ ), or fake news ( $p = 0.066$ ,  $CI = [0.055, 0.079]$ ). Morality plays a role in identifying categories by augmenting their probabilities of being classified as factual news, rumors, hate speech or junk science as the appeal to morality decreases; and on the contrary, it leads to higher probabilities of being classified as fake news clickbait or conspiracy theories as the appeal to morality in a given text increases.

## Discussion

The spread of misinformation content is an important and complex problem because not all misleading content is created equal. These results indicate that there are significant differences between content characteristics from factual sources and misinformation content in the form of clickbait, conspiracy theories, fake news, hate speech, junk science and rumors.

As for the size of these differences, I showed that fake news are, on average, 18 times more negative than factual news (having hate speech the highest difference with 30 times more negativity than factual news); that factual information is 15% more lexically diverse (with rumors being 19% less diverse than factual news); that fake news appeal to moral values 37% more than factual content (although hate speech appeals 50% more to morality than factual sources); and that fake news are 3% easier to process than factual sources (while fake news are 13% easier to process). From here, as seen in the hierarchical clustering, I showed how factual content is different from all the other categories and how fake news and conspiracy theories are closely related; that hate speech and rumors are also similar to conspiracy theories and hate speech or how clickbait and junk science are more similar to factual sources than to the other misinformation categories. These

results are important because the proposed fingerprints of misinformation provide grounding for the refinement of previous results and better interventions to mitigate their spread in the realm of technological companies, fact-checkers and media outlets or public policy in the light of the volume, breadth, speed and explainability challenges.

We know that our capacity for discerning between factual and false content is highly influenced by social norms (Frenda et al., 2011; Zhu et al., 2010) and moral foundations (Dehghani et al., 2016; Graham et al., 2009), defined as the “interlocking sets of values, practices, institutions, and evolved psychological mechanisms that function to suppress selfishness” (Graham et al., 2009, p. 1031). This is especially relevant for people with low levels of media literacy (Lazer et al., 2018) because the results show that, in general, misinformation requires less cognitive effort for being processed and is more reliant on emotions than factual information: it is easier to read, and falsified content creators use a less diverse vocabulary, which may explain why misinformation is more prevalent among low-literacy individuals (Lazer et al., 2018), and more appealing to negative emotions and moral values, which may influence our ability to discern between misinformation and factual content.

This has implications for previous and future research. It may be the case that the results moderate the motivated reasoning mechanism (Bago et al., 2020; Kraft et al., 2015) by augmenting or diminishing its effect depending on the level of emotional evocation and morality and the cognitive effort needed to process it. For example, emotions can be used as a strategy to convey a particular point of view when the appropriate methods are less likely to yield the desired conclusion (Kunda, 1990). Likewise, the results open the door to refining previous research. Preceding evidence shows that misinformation is more viral than factual news (Vosoughi et al., 2018). However, prior studies only focus on one or just a few specific categories, usually fake news. The obtained results offer the opportunity to explore to what extent different misinformation categories are more viral than others and how their structural features in terms of cognitive effort and appeal to emotions drive this virality.

In addition, this method can help technology companies, media outlets and fact-checking organizations to prioritize the content to be checked. This is aligned with the call for “public interest algorithms” to identify misinformation and protect consumers (Wheeler, 2017). With the model I provided and the resulting outcomes, one can inform more complex models to boost people’s abilities to identify misinformation. Regarding technology companies, there is evidence that correcting misinformation through “related stories” features in their platforms can reduce the prominence of misleading content (Bode & Vraga, 2015), although just around 3.7% (Pennycook & Rand, 2017). Beyond the small effect, these systems rely on content or sources that have already been flagged as potential misinformation. In contrast, using the results provided in this paper, these platforms may use the same system without the need to explicitly identify the source as misleading and recommending “related stories” based on the assessment of the readability, perplexity, emotionality, and morality levels of the content being shared. The feasibility of real-time deployment should not be a big problem since this method uses manageable calculations in a reasonable amount of time (less than one second for a 2000 words article). Therefore, the proposed method does not add friction to the user experience of a given platform. This will mitigate the volume, breadth, and speed challenges they face to mitigate the spread of misinformation, since the presented findings can be used to build machine-learning classification algorithms that could categorize content at the exact moment it is posted.

The findings also suggest promising interventions for early detection and the identification of check-worthy content, which is important for fact-checking organizations because the flood of misinformation can easily overwhelm them (Pennycook & Rand, 2017). The fact that factual content significantly differs in content characteristics from non-factual content opens the door for improving in the fact-checking process. Since it is impossible to check everything that is being shared on social media, fact-checkers focus on content that has already become viral (usually, with information provided from social media platforms in the form of rankings). However, the fingerprints can be used to automatically analyze every URL being shared in social media. If so, platforms and fact-checkers can complement their rankings based on virality with rankings based on the probability of a content being non-factual. More importantly, providing the differences of misinformation categories allows to focus on specific typologies, for example focusing on content targeting people's beliefs or influence major events (like fake news) and not so much on content to attract network traffic (like clickbait) (Vosoughi et al., 2018). Resembling the case of technology companies and social media platforms, the results contribute to mitigate the volume, breadth, and speed challenges that fact-checkers face when they manually assess content for veracity.

The results also highlight the importance of the efforts of public and private institutions to enhance media literacy (European Commission, 2018; Maksl et al., 2016). The notion of media literacy is based on the skills and competencies needed to successfully navigate a complex ecosystem (Eshet & Eshet, 2004; Guess et al., 2020). If people are aware that a specific content is aiming to trigger a reaction through less cognitive load and higher appealing to emotions, they will be less likely to share it (Martel et al., 2019). This is capital as our societies become more and more digital, especially in the light of worrying levels of media literacy. For example, the Pew Research Center found that only 17% of US adults have the skills to safely navigate the information ecosystem (Horrihan, 2016). However, investing in media literacy is particularly important in developing countries, where there are millions of Internet users with low-literacy rates (Mustaffa et al., 2007; Sun & Nekmat, 2008). In these situations, misinformation often escalates into violence, like in the case of ethnic brutality in India (McLaughlin, 2018) or violence against doctors in Latin America (Taylor, 2020). In similar contexts, improving media literacy to process more complex ideas will likely reduce the cognitive load associated with factual information (Ellis & Yuan, 2004; Swabey et al., 2016) and, therefore, the prevalence of misinformation. Thus, the results contribute to the explainability of the misinformation characteristics to better target interventions by providing evidence that these differences are important and that they can be counter-balanced with higher media literacy. For example, since my results indicate that misinformation is more emotional, media literacy efforts should include specific strategies to identify and raise awareness of highly emotional content.

However, this study also has some limitations. First, although I use a big sample of news articles, these are coming from a selection of websites. Even though I tried my best to select different sources for each category, I am aware of the potential limitation that the results could be based on domain-specific features. Secondly, while one of my research objectives was the explainability of the measures and models, I also acknowledge that—potentially—there may be more complex measures and models that may be more useful to discern between misinformation content and factual sources. The objective was not to provide the most detailed model to describe misinformation, but a parsimonious description of key features. Moreover, the results must be interpreted in the light of some methodological limitations that I expose in the corresponding section. Additionally,

although the objective of this paper was not discussing the evidence at the source level—but at the misinformation category level, it is also worth mentioning that sources belonging to the same category may behave differently. For example, some fake news sources may use a more neutral tone, while others may show a higher appeal to negative emotions. In other words, the results should be interpreted on the aggregated or average effect, and not in the specific source effect. Finally, although we employed widely used dictionaries in the literature, the results are also contingent to the specificities of each dictionary I employed.

Despite its limitations, the obtained results also suggest interesting paths for future research. For example, how is the fingerprints of misinformation evolving over time? Since misinformation creators are aware of counter-measure interventions, they may be adapting to them through adversarial machine learning (Huang et al., 2011), that is, changing their structural characteristics to avoid detection in automated systems. Similarly, it is interesting to seek if misinformation is becoming more disguised as factual information? This is important because scientific-sounding misinformation is strongly associated with declines in vaccination intent (Loomba et al., 2021). To do so, researchers could look at common characteristics and relationships between sources for each category. On the other hand, the transition from printed and televised news towards the digital world has caused a huge disruption in the traditional media industry, challenging their business models (Shearer, 2021). Focusing on the evolution of traditional media outlets could shed light on the question whether reliable sources are trying to draw on the mechanisms of virality to help increase their penetration (e.g., reliable news becoming more appealing to emotions, titles becoming more exciting, content easier to read, etc.)?

To the best of my knowledge, the findings presented in this paper are a compelling and impactful demonstration that not all misinformation is created equal. Of course, these differences have huge implications that also highlight the contributions. First, they contribute to diminish the problem of scalability in misinformation detection (the volume challenge), by providing a set of features that can differentiate between misinformation and factual content without human intervention. Secondly, they increase the evidence about structural differences among misinformation types (the breadth challenge), which allows to focus on or differentiate specific categories and create tailored countermeasures for each one. Thirdly, I propose a method that can be useful to identify misinformation before it becomes viral (the speed challenge), since the categorization of an URL can be done at the exact moment it is posted, allowing for on-spot detection. Finally, the model that is easily understandable in daily language (the explainability challenge), facilitating its adoption by technical and, specially, non-technical audiences.

### Data availability

The datasets generated during and/or analyzed during the current study are available in the following GitHub repository: <https://github.com/several27/FakeNewsCorpus>.

Received: 24 December 2021; Accepted: 21 April 2022;

Published online: 09 May 2022

### References

- Afroz S, Brennan M, Greenstadt R (2012) Detecting hoaxes, frauds, and deception in writing style online. *Proceedings-IEEE Symposium on Security and Privacy*. pp. 461–475. <https://ieeexplore.ieee.org/abstract/document/6234430>
- Akhtar MS, Ekbal A, Cambria E (2020) How intense are you? Predicting intensities of emotions and sentiments using stacked ensemble [application notes]. *IEEE Comput Intell Magaz* 15(1):64–75

- Alhabash S, Almutairi N, Lou C, Kim W (2019) Pathways to virality: psychophysiological responses preceding likes, shares, comments, and status updates on facebook. *Media Psychol* 22(2):196–216
- Allcott H, Gentzkow M, Yu C (2018) Trends in the diffusion of misinformation on social media. *Res Polit* 1–13. <https://doi.org/10.1177/2053168019848554>
- Appelman A, Sundar SS (2016) Measuring message credibility: construction and validation of an exclusive scale. *Journal Mass Commun Quart* 93(1):59–79
- Bago B, Rand DG, Pennycook G (2020). Fake news, fast and slow: deliberation reduces belief in false (but not true) news headlines. *J Exp Psychol Gen* 149:1–18
- Bai X (2011) Predicting consumer sentiments from online text. *Decision Supp Syst* 50(4):732–742
- Bakir V, McStay A (2018) Fake news and the economy of emotions: problems, causes, solutions. *Digit Journal* 6(2):154–175
- Bannard C, Lieven E, Tomasello M (2009) Modeling children's early grammatical knowledge. *Proc Natl Acad Sci USA* 106(41):17284–17289
- Barrón-Cedeño A, Jaradat I, Da San Martino G, Nakov P (2019) Propopy: organizing the news based on their propagandistic content. *Inform Process Manag* 56(5):1849–1864
- Beheshti Z, Nejadansari D, Barati H (2020) The relationship between emotional intelligence, Lexical diversity and the syntactic complexity of EFL Learners' written productions. *J Mod Res Engl Lang Stud* 7(1):133–161
- Berger J (2011) Arousal increases social transmission of information. *Psychol Sci* 22(7):891–893
- Berger JA, Milkman KL (2009) What makes online content viral? *Ssrn* 49(2):192–205
- Berger J, Milkman KL (2013) Emotion and virality: what makes online content go viral? *GfK Market Intell Rev* 5(1):18–23
- Bessi A, Coletto M, Davidescu GA, Scala A, Caldarelli G, Quattrociocchi W (2015) Science vs conspiracy: collective narratives in the age of misinformation. *PLoS ONE* 10(2):e0118093
- Bode L, Vraga EK (2015) In related news, that was wrong: the correction of misinformation through related stories functionality in social media. *J Commun* 65(4):619–638
- Bovet A, Makse HA (2019) Influence of fake news in Twitter during the 2016 US presidential election. *Nat Commun* 10(1):1–14
- Brady WJ, Crockett MJ, Van Bavel JJ (2020) The MAD model of moral contagion: the role of motivation, attention, and design in the spread of moralized content online. *Perspect Psychol Sci* 15(4):978–1010
- Brady WJ, Wills JA, Jost JT, Tucker JA, Van Bavel JJ, Fiske ST (2017) Emotion shapes the diffusion of moralized content in social networks. *Proc Natl Acad Sci USA* 114(28):7313–7318
- Broniatowski DA, Kerchner D, Farooq F, Huang X, Jamison AM, Dredze M, Ayers JW (2022) Twitter and Facebook posts about COVID-19 are less likely to spread misinformation compared to other health topics. *PLoS ONE* 17(1):e0261768
- Chakraborty K, Bhatia S, Bhattacharyya S, Platos J, Bag R, Hassanien AE (2020) Sentiment analysis of COVID-19 tweets by deep learning classifiers—a study to show how popularity is affecting accuracy in social media. *Appl Soft Comput* 97:106754
- Chan MpuiS, Jones CR, Hall Jamieson K, Albarracín D (2017) Debunking: a meta-analysis of the psychological efficacy of messages countering misinformation. *Psychol Sci* 28(11):1531–1546
- Choudhary A, Arora A (2021) Linguistic feature based learning model for fake news detection and classification. *Expert Syst Appl* 169(February 2020):114171
- Chung M, Kim N (2021) When I learn the news is false: how fact-checking information stems the spread of fake news via third-person perception. *Hum Commun Res* 47(1):1–24
- Cinelli M, Quattrociocchi W, Galeazzi A, Valensise CM, Brugnoli E, Schmidt AL, Scala A (2020) The COVID-19 social media infodemic. *Sci Rep* 2020 10:1 10(1):1–10
- Cox RH, Dickson D, Marier P (2020) Resistance, innovation, and improvisation: comparing the responses of nursing home workers to the COVID-19 pandemic in Canada and the United States. *J Comp Policy Anal Res Pract* 23(1):41–50
- D'Alessandro DM, Kingsley P, Johnson-West J (2001) The readability of pediatric patient education materials on the world wide web. *Arch Pediatr Adolesc Med* 155(7):807–812
- De Bruin WB, Granger Morgan M (2019) Reflections on an interdisciplinary collaboration to inform public understanding of climate change, mitigation, and impacts. *Proc Natl Acad Sci USA* 116(16):7676–7683
- de Souza JV, Gomes J, Souza Filho FMde, Oliveira Julio AMde, de Souza JF (2020) A systematic mapping on automatic classification of fake news in social media. *Soc Netw Anal Mining* 10(1):1–21
- Dehghani M, Johnson K, Hoover J, Sagi E, Garten J, Parmar N, ... Graham J (2016) Purity homophily in social networks. *J Exp Psychol Gen* 145(3):366–375
- Effron DA, Raj M (2020) Misinformation and morality: encountering fake-news headlines makes them seem less unethical to publish and share. *Psychol Sci* 31(1):75–87
- Ellis R, Yuan F (2004) The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Stud Second Lang Acquis* 26(1):59–84
- Eshet Y, Eshet Y (2004) Digital literacy: a conceptual framework for survival skills in the digital era. *J Educ Multimedia Hypermedia* 13(1):93–106
- European Commission (2018) Joint Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: Action Plan against Disinformation. High Representative of the Union for Foreign Affairs and Security Policy. European Commission
- Evans JSBT (2003) In two minds: dual-process accounts of reasoning. *Trend Cogn Sci* 7(10):454–459
- Fowler JH, Christakis NA (2009) Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study. *BMJ (Online)* 338(7685):23–26
- Frenda SJ, Nichols RM, Loftus EF (2011) Current issues and advances in misinformation research. *Curr Direc Psychol Sci* Vol. 20, pp. 20–23. SAGE Publications, Sage CA: Los Angeles, CA
- Fuller CM, Biros DP, Wilson RL (2009) Decision support for determining veracity via linguistic-based cues. *Decis Support Syst* 46(3):695–703
- Gamallo P, Campos JRP, Alegria I (2016) Comparing two basic methods for discriminating between similar languages and varieties\*. *Phys A: Stat Mech Appl* 170–177. <https://aanthology.org/W16-4822/>
- Gamallo P, Campos JRP, Alegria I (2017) A perplexity-based method for similar languages discrimination. Proceedings of the Fourth Workshop on NLP for Similar Languages, Varieties and Dialects. 109–114. <https://aanthology.org/W17-1213/>
- García-Portugés E (2021) Notes for predictive modeling Version 5.9.9. Available at <https://bookdown.org/egarpor/PM-UC3M/>
- Garrett RK, Weeks BE (2017) Epistemic beliefs' role in promoting misperceptions and conspiracist ideation. *PLoS ONE* 12(9):1–17
- Goel S, Anderson A, Hofman J, Watts DJ (2015) The structural virality of online diffusion. *Manag Sci* 1909:150722112809007
- González M (2015) An analysis of Twitter corpora and the differences between formal and colloquial tweets. In Proceedings of the Tweet Translation Workshop 2015 (pp. 1–7)
- Graham J, Haidt J, Nosek BA (2009) Liberals and conservatives rely on different sets of moral foundations. *J Person Soc Psychol* 96(5):1029–1046
- Griffiths TL, Steyvers M (2004) Finding scientific topics. *Proc Natl Acad Sci USA* 101(suppl 1):5228–5235
- Grinberg N, Joseph K, Friedland L, Swire-Thompson B, Lazer D (2019) Fake news on Twitter during the 2016 U.S. presidential election. *Science* 363(6425):374–378
- Guess AM, Lerner M, Lyons B, Montgomery JM, Nyhan B, Reifler J, Sircar N (2020) A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proc Natl Acad Sci USA* 117(27):15536–15545
- Hakak S, Alazab M, Khan S, Gadekallu TR, Maddikunta PKR, Khan WZ (2021) An ensemble machine learning approach through effective feature extraction to classify fake news. *Future Gener Comput Syst* 117:47–58
- Hansen LK, Arvidsson A, Nielsen FA, Colleoni E, Etter M (2011) Good friends, bad news-affect and virality in twitter. *Commun Comput Inf Sci* 185 CCIS(PART 2):34–43
- Hartiga J, Wong M (1979) A K-means clustering algorithm. *Appl Stat* 28:126–130
- Hauch V, Blandón-Gitlin I, Masip J, Sporer SL (2015) Are computers effective lie detectors? A meta-analysis of linguistic cues to deception. *Person Soc Psychol Rev* 19(4):307–342
- Hee CVan, Jacobs G, Emmery C, Desmet B, Lefever E, Verhoeven B, Hoste V (2018) Automatic detection of cyberbullying in social media text. *PLoS ONE* 13(10):e0203794
- Helmstetter S, Paulheim H (2018) Weakly supervised learning for fake news detection on Twitter. Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2018, pp. 274–277
- Horberg EJ, Oveis C, Keltner D (2011) Emotions as moral amplifiers: An appraisal tendency approach to the influences of distinct emotions upon moral judgment. *Emot Rev* 3(3):237–244
- Horrigan J (2016) Digital literacy and learning in the United States
- Huang L, Joseph AD, Nelson B, Rubinstein BIP, Tygar JD (2011) Adversarial machine learning. Proceedings of the 4th ACM Workshop on Security and Artificial Intelligence. Association for Computing Machinery, New York, NY, USA
- Ju Y, Back KJ, Choi Y, Lee JS (2019) Exploring Airbnb service quality attributes and their asymmetric effects on customer satisfaction. *Int J Hosp Manag* 77:342–352



- Just MA, Carpenter PA (1980) A theory of reading: from eye fixations to comprehension. *Psychol Rev* 87(4):329–354
- Kietzmann J, Lee LW, McCarthy IP, Kietzmann TC (2020) Deepfakes: trick or treat? *Bus Horiz* 63(2):135–146
- Kincaid JP, Fishburne Jr, RP, Rogers RL, Chissom BS (1975) Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Naval Technical Training Command Millington TN Research Branch
- King G, Tomz M, Wittenberg J (2000) Making the most of statistical analyses: improving interpretation and presentation. *Am J Polit Sci* 44(2):347
- Koenecke A, Nam A, Lake E, Nudell J, Quartey M, Mengesha Z, Goel S (2020) Racial disparities in automated speech recognition. *Proc Natl Acad Sci USA* 117(14):7684–7689
- Kononova A, Joo E, Yuan S (2016) If I choose when to switch: Heavy multitaskers remember online content better than light multitaskers when they have the freedom to multitask. *Comput Hum Behav* 65:567–575
- Kraft PW, Krupnikov Y, Milita K, Ryan JB, Soroka S (2020) Social media and the changing information environment: sentiment differences in read versus recirculated news content. *Public Opin Quart* 84(S1):195–215
- Kraft PW, Lodge M, Taber CS (2015) Why people “Don’t Trust the Evidence”: motivated reasoning and scientific beliefs. *Ann Am Acad Polit Soc Sci* 658(1):121–133
- Kramer ADI, Guillory JE, Hancock JT (2014) Experimental evidence of massive-scale emotional contagion through social networks. *Proc Natl Acad Sci USA* 111(24):8788–8790
- Kulkarni KK, Kalro AD, Sharma D, Sharma P (2020) A typology of viral ad sharers using sentiment analysis. *J Retail Consum Serv* 53:101739
- Kunda Z (1990) The case for motivated reasoning. *Psychol Bull* 108(3):480–498
- Lang A (2000) The limited capacity model of mediated message processing. *J Commun* 50(1):46–70
- Lang A (2006) Using the limited capacity model of motivated mediated message processing to design effective cancer communication messages. *J Commun* 56(SUPPL):57–80
- Lazer DMJ, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, Zittrain JL (2018) The science of fake news. *Science* 359(6380):1094–1096
- Leshner G, Cheng IH (2009) The effects of frame, appeal, and outcome extremity of antimoking messages on cognitive processing. *Health Commun* 24(3):219–227
- Leshner G, Vultee F, Bolls PD, Moore J (2010) When a fear appeal isn’t just a fear appeal: the effects of graphic anti-tobacco messages. *J Broadcast Electron Media* 54(3):485–507
- Lewandowsky S, Ecker UKH, Seifert CM, Schwarz N, Cook J (2012) Misinformation and its correction: continued influence and successful debiasing. *Psychol Scie Public Interest, Suppl* 13(3):106–131
- Liu Z, Park S (2015) What makes a useful online review? Implication for travel product websites. *Tour Manag* 47:140–151
- Loomba S, de Figueiredo A, Piatek SJ, de Graaf K, Larson HJ (2021) Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nat Hum Behav* 5(3):337–348
- Mahir EM, Akhter S, Huq MR (2019). Detecting fake news using machine learning and deep learning algorithms. In 2019 7th International Conference on Smart Computing & Communications (ICSCC) (pp. 1–5). IEEE. Miri, Sarawak, Malaysia
- Maksl A, Craft S, Ashley S, Miller, D (2016) The usefulness of a news media literacy measure in evaluating a news literacy curriculum. 72(2), 228–241. <https://doi.org/10.1177/10776958166651970>
- Manzoor IS, Singla J, Nikita (2019) Fake news detection using machine learning approaches: a systematic review. Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019)
- Martel C, Pennycook G, Rand DG (2019) Reliance on emotion promotes belief in fake news. *Cogn Res Princ Implic*. 5:47
- Masciari E, Moscato V, Picariello A, Sperli G(2020) A deep learning approach to fake news detection. Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics), 12117 LNAI 3:113–122
- Matalon Y, Magdaci O, Almozilino A, Yamin D (2021) Using sentiment analysis to predict opinion inversion in Tweets of political communication. *Sci Rep* 11(1):1–9
- McCornack SA, Morrison K, Paik JE, Wisner AM, Zhu X (2014) Information manipulation theory 2: a propositional theory of deceptive discourse production. *J Lang Soc Psychol* 33(4):348–377
- McLaughlin T (2018) How whatsapp fuels fake news and violence in India. WIRED <https://www.wired.com/story/how-whatsapp-fuels-fake-news-and-violence-in-india/>
- Milkman KL, Berger J (2014) The science of sharing and the sharing of science. *Proc Natl Acad Sci USA* 111:13642–13649
- Miller T, Howe P, Sonenberg L (2017) Explainable AI: beware of inmates running the asylum or: how i learnt to stop worrying and love the social and behavioural sciences. <https://arxiv.org/abs/1712.00547>
- Mocanu D, Rossi L, Zhang Q, Karsai M, Quattrociocchi W (2015) Collective attention in the age of (mis)information. *Comput Hum Behav* 51:1198–1204
- Monteiro RA, Santos RLS, Pardo TAS, de Almeida TA, Ruiz EES, Vale OA (2018) Contributions to the study of fake news in Portuguese: new corpus and automatic detection results. International Conference on Computational Processing of the Portuguese Language. December. pp. 324–334
- Mustaffa A, Norsiah A, Mustaffa C (2007) Media literacy: accessibility and skills among Malaysian women. *Intercult Commun Stud* XVI: 3
- Neuman WR, Just MR, Crigler AN (1992) Common knowledge: News and the construction of political meaning. University of Chicago Press, Chicago
- Nielsen FÅ (2011) A new ANEW: evaluation of a word list for sentiment analysis in microblogs. *CEUR Workshop Proc* 718:93–98
- Pennycook G, Rand D (2017) The implied truth effect: attaching warnings to a subset of fake news stories increases perceived accuracy of stories without warnings. *Manag Sci* 66:11
- Pennycook Gordon, Allan J, Nathaniel C, Derek B, Fugelsang KJA (2015) On the reception and detection of pseudo-profound bullshit. *Judgem Decis Making* 10(6):549–563
- Pennycook Gordon, Rand DG (2019) Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition* 188(June 2018):39–50
- Qiu G, He X, Zhang F, Shi Y, Bu J, Chen C (2010) DASA: Dissatisfaction-oriented advertising based on sentiment analysis. *Exp Syst Appl* 37(9):6182–6191
- Ragini JR, Anand PMR, Bhaskar V (2018) Big data analytics for disaster response and recovery through sentiment analysis. *Int J Inform Manag* 42:13–24
- Rajadesingan A, Zafarani R, Liu H (2015) Sarcasm detection on Twitter: A behavioral modeling approach. In Proceedings of the eighth ACM international conference on web search and data mining (pp. 97–106). ACM Association for Computing Machinery
- Rashkin H, Choi E, Jang JY, Volkova S, Choi Y (2017) Truth of varying shades: Analyzing language in fake news and political fact-checking. In Proceedings of the 2017 conference on empirical methods in natural language processing (pp. 2931–2937). ACL Association for Computational Linguistics
- Rathje S, Van Bavel JJ, van der Linden S (2021) Out-group animosity drives engagement on social media. *Proc Natl Acad Sci USA* 118(26):e2024292118
- Risdal M (2016) Getting real about fake news | Kaggle
- Rosenquist JN, Fowler JH, Christakis NA (2011) Social network determinants of depression. *Mol Psychiatr* 16(3):273–281
- Rubin VL, Chen Y, Conroy NK (2015) Deception detection for news: Three types of fakes. *Proc Assoc Inform Sci Technol* 52(1):1–4
- Sagi E, Dehghani M (2014) Measuring moral rhetoric in text. *Soc Sci Comput Rev* 32(2):132–144
- Scheufele DA, Krause NM (2019) Science audiences, misinformation, and fake news. *Proc Natl Acad Sci USA* 116(16):7662–7669
- Shearer E (2021) More than eight-in-ten Americans get news from digital devices. Pew Research Center, 12
- Singh L, Bode L, Budak C, Kawntiranon K, Padden C, Vraga E (2020) Understanding high- and low-quality URL Sharing on COVID-19 Twitter streams. *J Comput Soc Sci* 3(2):343–366
- Stanovich, KE (2005) The robot’s rebellion: finding meaning in the age of Darwin. University of Chicago Press, Chicago, IL
- Sun S, Nekmat E (2008) Learning through “prosuming”: insights from media literacy programmes in Asia. *Technol Soc* 13:259–278
- Swabey L, Nicodemus B, Taylor MM, Gile D (2016) Lexical decisions and related cognitive issues in spoken and signed language interpreting. *Int J Res Pract Interpret* 18(1):34–56
- Szpakowski M (2018) Fake news corpus. [https://github.com/several\\_27](https://github.com/several_27)
- Taddicken M, Wolff L (2020) ‘Fake News’ in science communication: emotions and strategies of coping with dissonance online. *Media Commun* 8(1):206–217
- Tajfel H, Turner JC, Austin WG, Worchel, S (1979) An integrative theory of intergroup conflict. *Organizational identity: A reader*, 56(65), 9780203505984–16
- Tambuscio M, Oliveira DFM, Ciampaglia GL, Ruffo G (2018) Network segregation in a model of misinformation and fact-checking. *J Comput Soc Sci* 1(2):261–275
- Tambuscio M, Ruffo G, Flammini A, Menczer F (2015) Fact-checking effect on viral hoaxes: A model of misinformation spread in social networks. In Proceedings of the 24th international conference on World Wide Web (pp. 977–982). ACM Association for Computing Machinery
- Taylor L (2020) Covid-19 misinformation sparks threats and violence against doctors in Latin America. *BMJ* 370 (pp. 1–2)
- Tsugawa S, Ohsaki H (2017) On the relation between message sentiment and its virality on social media. *Soc Netw Anal Mining* 7(1):1–14.
- Tumasjan A, Sprenger T, Sandner P, Welpe I (2010) Predicting elections with Twitter: what 140 characters reveal about political sentiment. Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, 4(1). Washington, DC, USA, May 23–26, 2010

- Turner J, Hogg M, Oakes P, Reicher S, Wetherell M (1987) Rediscovering the social group: a self-categorization theory. *PsycNET*. Basil Blackwell
- van Dijck J (2013) “You have one identity”: performing the self on Facebook and LinkedIn. *Media Cult Soc* 35(2):199–215
- Venables W, Ripley B (2002) *Modern applied statistics with S* (Fourth Edn). Springer, New York
- Vosoughi S, Roy D, Aral S (2018) The spread of true and false news online. *Science* 359(6380):1146–1151
- Wang WY (2017) “Liar, liar pants on fire”: a new benchmark dataset for fake news detection. *ACL 2017-55th Annual Meeting of the Association for Computational Linguistics. Proc Conference (Long Papers)* 2:422–426
- Weeks BE (2015) Emotions, partisanship, and misperceptions: how anger and anxiety moderate the effect of partisan bias on susceptibility to political misinformation. *J Commun* 65(4):699–719
- Wheeler T (2017) Using “public interest algorithms” to tackle the problems created by social media algorithms
- Zervopoulos A, Alvanou AG, Bezas K, Papamichail A, Maragoudakis M, Kermanidis K (2020) Hong Kong protests: using natural language processing for fake news detection on twitter. *IFIP Advances in Information and Communication Technology*, 584 IFIP, Springer, pp. 408–419
- Zhou X, Zafarani R (2020) A survey of fake news: fundamental theories, detection methods, and opportunities. *ACM Computing Surveys*, 53(5) <https://arxiv.org/abs/1812.00315>
- Zhu B, Chen C, Loftus EF, Lin C, He Q, Chen C, Dong Q (2010) Individual differences in false memory from misinformation: Personality characteristics and their interactions with cognitive abilities. *Pers Individ Differ* 48(8):889–894
- Zimdars, M (2017) *OpenSources*. Professionally curated lists of online sources, available free for public use. <https://github.com/BigMcLargeHuge/opensources>
- Zuckerman M, DePaulo B, Rosenthal R (1981) Verbal and nonverbal communication of deception. *Adv Exp Soc Psychol* 14:1–59

### Acknowledgements

Part of this research has been financed by the Spanish Ministry of Science and Innovation and the European Social Fund PRE2019-091668.

### Competing interests

The author declares no competing interests.

### Ethical approval

This article does not contain any studies with human participants performed by the author.

### Informed consent

Does not apply.

### Additional information

**Correspondence** and requests for materials should be addressed to Carlos Carrasco-Farré.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

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



**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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022