Developing fake news immunity: fallacies as misinformation triggers during the pandemic

ABSTRACT

Misinformation constitutes one of the main challenges to counter the infodemic: misleading news, even if not blatantly false, can cause harm especially in crisis scenarios such as the pandemic. Due to the fast proliferation of information across digital media, human fact-checkers struggle to keep up with fake news, while automatic factcheckers are not able to identify the grey area of misinformation. We, thus, propose to reverse engineer the manipulation of information offering citizens the means to become their own fact-checkers through digital literacy and critical thinking. Through a corpus analysis of fact-checked news about COVID-19, we identify 10 fallacies – arguments which seem valid but are not – that systematically trigger misinformation and offer a systematic procedure to identify them. Next to fallacies, we observe the types of sources associated to (mis-/dis-)information in our dataset as well as the type of claims making up the headlines. The observation of these three levels of analysis reveals a misinformation ecosystem where developing the audience's digital literacy is necessary to guarantee fake news immunity.

Keywords: misinformation, fallacy theory, digital literacy, fact checking, multi-level annotation

[Click here to download the Word file]

21 **Comments from Reviewer1** (changes highlighted in yellow)

22 **B. Article presentation:**

23 1.   According to "Author Guidelines" which is shown in the OJCMT web
24 address https://www.ojcmt.net/home/author-guidelines , author(s) should consult the APA Style
25 Manual (7th Edition). Please make sure to use the APA7 style for your whole manuscript. You can also
26 see the reviewer's comments on your manuscript body file. Please double-check especially :

27   a.   Please consider APA7 In-text citations style for the whole manuscript.   For
28   APA style sample papers , you can visit : https://apastyle.apa.org/style-grammar-
29   guidelines/paper-format/professional-annotated.pdf

30 Amended in line with APA7 guidelines

31   b.   Missing punctuations for the text body and in-text citations. For further details please
32   see Rewierer's comments on embedded document.

33 Amended in line with reviewer's comments

34   c.   Please make sure to use your narrative citations correctly. There are numerous errors
35   of using " Narrative In-Text Citations". For example lines 71-72, lines 426-432...Please see
36   tracked document for further information.

37   https://apastyle.apa.org/style-grammar-guidelines/citations/basic-principles/parenthetical-
38   versus-narrative

39 Amended in line with APA7 guidelines

40   d.   Please ensure that to use scientific language instead of colloquial language in
41   sentences. (For example, line 172, line 187, see also Reviewer's comments)

42 Amended in line with APA7 guidelines

43   e.   Do not add extra space before or after paragraphs. https://apastyle.apa.org/style-
44   grammar-guidelines/paper-format/line-spacing

45 Amended in line with APA7 guidelines

46   d.   Please check once again your manuscript about passionately using two em dashes -
47   ....- and quotation marks ".....".

There are also lots of incorrect usages in using two em dashes and quotation marks together (e.g. lines 46-48, see also Reviewer's comments). There are many incorrect usages in using two em dashes.

Amended in line with APA7 guidelines

e. Using two em-dashes, please make sure to use the correct format. (Line 409-411,..) Please see also Reviewer's comments

Amended in line with Reviewer's comments

f. Please use lowercase letters to separate items. (Line 205-206, for more correction please see reviewer's comments) https://apastyle.apa.org/style-grammar-guidelines/lists/lettered

 Amended in line with APA7 guidelines

g. There are too many "two different usages" of the same terms (such as Covid-10/COVID-19) and punctuation marks (such as em dashes and quotation marks). Please make sure that you standardised your wording and your grammar. (For full corrections please see also Reviewer's comments embedded document) .

we have opted for a uniform Covid-19 in line with the majority of scientific articles on the matter (https://www.theguardian.com/commentisfree/2020/apr/19/covid-pedantry-national-crisis-spelling-grammar)

h. While using website links in your text, do not create references or in-text citations for whole websites (For example line 431-432, for more correction please see reviewer's comments). For more information also please visithttps://apastyle.apa.org/style-grammar-guidelines/references/examples/whole-website-references . If you want to use specific website data, you can prefer footnotes.

ı. For tables and figures, please make sure to follow "Author Guideline".

Following the Guidelines, figures and table have been re-inserted in the body of text with the required formatting

j. Presentation of research questions: Please clarify and highlight your research questions in your body text. RQ1, RQ2...

Research questions have been slightly reformulated to clarify their scope and they have been matched with results sections and highlighted in the conclusion to signpost how they have been answered through the analysis

78   k. Seperation of parallel items: There are several usage in you manuscript (lines 95-96, lines

79   205-206, lines 241-242, lines 369-370) . Authors should use lowercase letters in parentheses

80   before each item.

81   Amended

82   2. Grammar and scientific language

83   a. Terms and symbols: If you use a specific term for a special purpose, please explain. For

84   example "Golden annotation" cannot be found in the relevant literature. Also please make an

85   explanation of using [p]eople instead of people.

86   Golden annotation is a technical term employed by the Natural Language Processing

87   community. The meaning of the term has been clarified within the text and replaced with

88   "golden standard annotation" to make it more transparent for the wider audience.

89   [p]eople replaced with people

90   b. Grammar and spelling: There are some grammar and spelling corrections, please see

91   reviewer's comments)

92   Checked and revised according to reviewer's comments.

93   c. Scientific language: Please make sure to use scientific language in your manuscript. "Let

94   us" , " a piece of fake" etc...

95   amended

96 **C. Methodology and scientific soundness:**

97   1.   Manuscript Abstract: Please check the balance of "purpose, methodology, and

98   results" sections in your abstract. More importantly, your result section is missing.

99   The results have been summarized in the abstract and balanced with the other sections

100

101

102   2.   Manuscript body

103   a.   Theoretical framework: Please strengthen your theoretical background. The

104   theoretical section lacks literature, please be sure to add more literature instead of

105   an introduction to digital literacy.

106   <mark>The theoretical framework has been expanded with a focus on the notion of digital</mark>

107   <mark>media literacy and its connection with critical thinking that underpins the relevance</mark>

108   <mark>of fallacy analysis.</mark>

109

110   b. Missing data: Assessment cannot be made in the analysis section due to missing

111   tables and figures. For tables and figures please follow "Author Guideline".

112   <mark>Tables and figures have been re-inserted</mark>

113

114   c. Conclusion with a question: Please ensure the pupose of "conclusion section" in a

115   scientific article. (The question in between lines 616 and 620)

116   <mark>Questions have been replaced with affirmative statements</mark>

117

118   OJCMT-13854-2021-R1-TRACKED.asd.docx

119

120   **Comments from Reviewer 2**

121   The article is very carefully written. Authors offer a compact theoretical chapter, which substantiates the essence

122   of the research project. It has a very detailed method description section. The article makes it easy to keep track

123   of how the method was used, what problems were solved during the study, while both collecting data and

124   improving the chosen research method. In the course of the research, both qualitative text analysis methods and

125   quantitative analysis of the obtained data were used. The transparency of the research process makes it possible

126   to assess the quality of the research and the scientific reliability of the data. The development of the fallacies

127   decatlon proposed in the article makes this study results usable for practical methods development in fact-

128   finding and media organizations.

129   no changes required

130   **Comments from Reviewer 3**

131   The article lacks a methodology. There should be a clear methodology.

132   The article, as highlighted by Reviewer 2, has a clear methodology which encompasses a i)

133   systematic dataset collection through web scraping ii) a multilevel corpus annotation with

134   multiple annotators verified through Inter-Annotator agreement metrics and confusion

135   matrices based on theoretically informed guidelines iii) the inter and intra analysis of the

136   results looking at frequency distributions and statistical significance. The methodological
137   pipeline has been summarized in the introduction (text highlighted in green).
138   <mark>In view of the second round of review, we have explicitly highlighted the methodology</mark>
139   <mark>section</mark>
140

141   # Developing fake news immunity: fallacies as misinformation
142   # triggers during the pandemic

143   **Abstract:**
144   Misinformation constitutes one of the main challenges to counter the infodemic: misleading news,
145   even if not blatantly false, can cause harm especially in crisis scenarios such as the pandemic. Due
146   to the fast proliferation of information across digital media, human fact-checkers struggle to keep
147   up with fake news, while automatic fact-checkers are not able to identify the grey area of
148   misinformation.We, thus, propose to reverse engineer the manipulation of information offering
149   citizens the means to become their own fact-checkers through digital literacy and critical thinking.
150   Through a corpus analysis of fact-checked news about Covid-19, we identify 10 fallacies –
151   arguments which seem valid but are not – that systematically trigger misinformation and offer a
152   systematic procedure to identify them. Next to fallacies, we examine the types of sources
153   associated to (mis-/dis-)information in our dataset as well as the type of claims making up the
154   headlines. <mark>The statistical patterns surfaced from these three levels of analysis reveal a</mark>
155   <mark>misinformation ecosystem where no source type is exempt from flawed arguments with frequent</mark>
156   <mark>evading the burden of proof and cherry picking behaviors, even when descriptive claims are at</mark>
157   <mark>stake. In such a scenario, exercising the audience's critical skills through fallacy and semantic</mark>
158   <mark>analysis is necessary to guarantee fake news immunity.</mark>
159

160   **Keywords:** misinformation, fallacy theory, digital literacy, fact-checking, multi-level annotation
161

162   [Click here to download the Word file]
163

INTRODUCTION

One of the major challenges of the current information ecosystem is the rapid spread of misinformation through digital media. Initial discussions of infodemiology – the role of information spread in support of or exacerbating issue of health and health policy – has brought to the fore the need to improve fact-checking to counter intentional and unintentional misbehaviours and inform policy making. The buzzword "fake news" has been used to refer to phenomena ranging from news, parody, to propaganda, and news fabrication. Even when adopting a strict definition of fake news as intentionally lacking facticity to a certain degree (Tandoc et al., 2018), there are clear variations: a news claiming that "mRNA vaccines are capable of altering or damaging human DNA" (Kasprak, 2020) is more fake than a news claiming that "Vaccines are unavoidably unsafe" (Teoh, 2020). Both might trigger wrong perceptions and attitudes, but the latter news claim does not convey entirely false information. As explained by the fact-checker organization *Healthfeedback.org*, the legal phrase "unavoidably unsafe", which takes into account risk/benefit trade-offs, leads to misleading interpretations of the vaccine as "dangerous".

Due to continuous updates about Covid-19 from the scientific community as well as governments and health institutions, the media may unintentionally disseminate misleading content which goes beyond lexical vagueness by, for example, drawing defeasible generalizations out of partial scientific results or single anecdotes (Molina et al., 2019). In other words, what makes these types of news *fake* is not just the truth of the information conveyed. Rather it is the misleading presentation or reasoning of the arguments they convey. This is done, for example, through false analogies, hasty generalizations, and cherry picking of information. This type of fake news is generally addressed as *misinformation,* which is the distribution of information which is not necessarily false and not deliberately created to harm (Yates et al., 2020). Even though unintentionally dangerous, misinformation has a wide societal impact. Brennen et al. (2020) found that 59% of fake news does not contain either fabricated or imposter content, but rather reconfigured misinformation. This misinformation proliferates through social media, the main source of news for infodemically vulnerable citizens. In other research (Carmi et al., 2020), limited types of source and information checking across both social media and search engine and reliance on the opinions of close friends and family have been identified as corresponding to low levels of digital and data literacy.

However, the identification of misinformation is far from being successfully addressed by human fact-checkers, let alone automated ones. The rating categories of different fact-checker organizations represented in the Google Fact Checker initiative lack of an agreed truth barometer based on systematic, mutually exclusive and clear criteria, thus hindering public understanding. As a result, datasets coded as misinformation that can be used to train systems for automatic fact-checking of information are scarce, even though needed. As remarked by Thorne and Vlachos (2018), current text classification approaches leveraging fact-checked datasets of claims are not enough since additional contextual information alongside factuality is required to capture misinformation.

To lay the foundations for a fact-checking process that uncovers misinformation triggers, we propose a systematic and multilevel procedure to identify fallacious arguments. Our theoretical assumption is that fallacies, arguments that seem valid but are not, work as indicators of misinformation. We apply our system to

the analysis of a dataset of 1135 Covid-19 related fact checked news, revealing major trends in the way misinformation is constructed and communicated. From an empirical perspective, we adopt a bottom-up approach focusing on the specific characteristics of the news reports: that is, we develop a set of guidelines for the identification of: a) fallacies (e.g. false authority); b) the type of media source hosting the news (e.g. social media; broadcast digital news) and c); the semantic type of claim expressing the news title (e.g. prediction vs. interpretation). We conduct an annotation experiment with two non-expert annotators and then check disagreement cases emerging from the inter-annotator agreement metrics through the aid of an expert annotator (golden standard annotation). We then focus on statistical trends which feature the golden standard annotations looking at the frequency of values for each analytic level as well as $\chi 2$ contingency tables across different levels of analysis to answer the following questions:

RQ1. Is it possible to develop a reliable procedure for the identification of misinformation triggers?

RQ2. What are the triggers of misinformation (fallacies, types of claims) and what is their frequency?

RQ3. What sources are more likely to spread misinformation?

RQ4. Do certain sources tend to be associated with certain fallacies and/or types of claims and vice versa?

In the following sections, we show how we developed our theoretical approach based on Fallacy Theory (Hamblin, 1970). We then move onto explaining how we designed the classification system of most common fallacies in news relating to Covid-19. After that, we zoom into the categories that have triggered agreement and disagreement among the annotators. We then move to the results of the analysis we conducted of the news articles pointing to statistically significant trends. Finally, we discuss our findings and how they can contribute to a better understanding how to educate society about online news manipulations.

Theoretical framework

Digital Media literacy to fight the infodemic

The Covid-19 pandemic meant that millions of people across the world were moving in and out of lockdowns and had to rely on digital systems and news sites for their everyday needs. But beyond digital divides around access to the Internet there is also the issue of digital media literacy. For example, Abdul-Fatawu Abdulai et al. (2020) have examined Covid-19 related digital skills among people in Ghana and argue that people experienced challenges in locating the appropriate online resources related to the pandemic. Importantly, they found that people experienced difficulties in distinguishing good quality information from opinions and anecdotes. Similarly, Beaunoyer, Dupere, and Guitton (2020) argue that people who have lower digital health related skills are more vulnerable to getting infected and infecting others because they have more challenges in accessing, understanding and applying the proper measures. As they argue, "people not able to decipher the degree of veracity of information (typically due to low level of critical digital or health literacy) might follow various advice regarding Covid-19 that could not only be detrimental for their health but also be harmful for the population" (Beaunoyer et al., 2020). One of the avenues they propose to mitigate digital inequalities related to Covid-19 is to improve people's ability to detect fake-news.

According to Fletcher et al. (2020), in the UK there was an interest in news in the beginning of the pandemic that slowly decreased. However, access to news about Covid-19 was unevenly distributed, with people who come from lower socio-economic status in terms of levels of education (this factor is especially dominant in

241 online news consumption) and household income being less likely to consume news. As Fletcher et al. (2020)

242 identified throughout the pandemic, people used social media in high proportions but as time progressed the use

243 of social media for news and information about Covid-19 decreased. Nevertheless, the proportion of people who

244 say they avoid news increased to 25% in early June 2020, a trend that is influenced by various factors such as

245 the negative effect on mood.

246 In relation to engagement with fake news and misinformation, Kyriakidou et al. (2020) argue that UK citizens

247 "felt misled by a range of information they encountered, which – in their view – was often conflicted or

248 inconclusive, including government claims about the human impact of the pandemic in the UK". According to

249 them, people felt that the most confusing and misleading content they encountered came from the UK

250 government's messages during the pandemic. In this context, some scholars (Amazeem and Bucy, 2019; Kahne

251 and Bowyer, 2017; Vraga et al., 2020) argue that teaching people news literacy might be one solution. News

252 literacy is defined as having an understanding about the processes of producing, distributing, and engaging with

253 news. More specifically, news literacy can "provide a foundation to improve information consumption processes

254 by giving social media users the tools to identify, consume, and share high-quality information regarding Covid-

255 19" (Vraga et al., 2020). In the era of Networked society, to be able to responsibly consume and produce news

256 implies being a media literate person who "can decode, evaluate, analyze and produce both print and electronic

257 media. The fundamental objective of media literacy is a critical autonomy relationship to all media"

258 (Aufderheide, 1992). The centrality of media literacy to counter fake news has been recently underlined by the

259 European Commission in their Action plan against disinformation as requiring "continuous and sustained efforts

260 to support education and media literacy, journalism, fact-checkers, researchers, and the civil society as a whole"

261 (2018).[1] Scholars have repeatedly pointed to critical thinking as the kernel of media literacy. Hobbs (2011)

262 considers, for instance, "Comprehending messages and using critical thinking to analyze message quality,

263 veracity, credibility […]" as the second component (access; analyse and evaluate; create; reflect; act) of the five

264 essential to develop media literacy. Similarly, Koltay (2011) defines "Having a critical approach to quality and

265 accuracy of content" among the five stages to build media literacy. However, so far, no systematic intervention

266 to teach critical thinking in the news environment has been carried out. The ability to evaluate whether the

267 arguments that form news are correct or fallacious contributes to this endeavor, constituting part and parcel of

268 the critical thinking needed to be a digital media literate.

269 **Rhetorical clues (fallacies) to identify misinformation**

270 The theoretical basis of our approach is founded on the notion of fallacy. A standard definition of fallacy that

271 goes back to Aristotle is an argument that "seems to be valid but is not so" (Hamblin, 1970; Tindale, 2007).

272 Aristotle has undoubtedly provided the foundations for the systematic study of fallacious arguments,[2] even if the

273 'textbook' versions he neatly outlines may be rarely found in real life discourse. Because fallacious arguments

274 can be very close to valid ones sometimes it may be difficult to talk about clear-cut distinctions (Boudry et al.,

---

[1] Action plan against disinformation. "Joint Communication The European Parliament, The European Council, The Council, The European Economic And Social Committee And The Committee Of The Regions." (2018). https://ec. europa. eu/newsroom/dae/document. cfm

[2] *Sophistical Refutations*; *Rhetoric* 2.24. The first theoretical discussions of fallacious reasoning can be traced back to Gorgias (now mostly lost and fragmented) and Plato (e.g., *Hippias Minor*, *Euthydemus*).

275 2015). More significantly, this closeness explains why fallacious reasoning is persuasive because it follows,

276 even if partially, the patterns and tropes of non-fallacious reasoning thereby producing arguments that are not

277 entirely invalid or outrageously unacceptable, at least at first glance. With a focus to the realm of mis-/dis-

278 /information, the persuasiveness of misinformation can be explained in a similar vein: 'fake news' can be

279 viewed as news that 'seems to be valid but is not so'. For example, the fallacy of 'cherry-picking' may happen

280 intentionally or unintentionally when specific information that supports a given position is chosen, while

281 ignoring or dismissing information which does not support it.[3] This means that an instance of fake news that is

282 the outcome of cherry-picking can be based on *partial* information, but not necessarily *false* information. Such

283 combinations of valid and invalid information, and arguments that we often encounter in discourse that involves

284 fallacious reasoning, shows why misinformation has a grip on people. This becomes even more evident when

285 we turn to news in the realm of misinformation. Fallacy identification is an efficient way for achieving bottom-

286 up deconstruction of misinformation that privileges misinformation 'pre-bunking' over debunking. While fact-

287 checking websites attempt to categorize misinformation on the basis of truth barometers that are partially

288 informative (e.g. labels such as "half true"), fallacy identification points directly to the roots of the

289 misinformation problem. In particular, fallacy identification copes with the grey areas of misinformation and

290 allows us to draw and analyze its different shades in a qualitative and constructive way that could never be

291 achieved through the available truth barometers. Importantly, it helps us learn how to identify misinformation

292 and cope with online manipulations.

293 The relevance of fallacies can be showcased through the analysis of a case from our Covid-19 news dataset. A

294 claim circulated on Facebook that 'the flu shot causes false positive results on Covid-19 tests' has been fact-

295 checked by *Healthfeedback.org* and assigned the label "Incorrect". How helpful, however, is the label

296 "Incorrect" or, to take a few more from the same truth barometer, "Misleading", "Half True", and "Inaccurate"

297 for evaluating and deconstructing misinformation? Such labels merely indicate that there is something flawed

298 with the news at hand, but they do not provide constructive insights about the nature of misinformation.

299 Fallacy identification, on the other hand, explains the roots of misinformation, whether it relies on quantity and

300 quality of evidence available, the type of reasoning at stake or the language involved. In the case at hand, for

301 example, the dominant fallacy at stake is that of *post hoc*: the fact that coronavirus was detected in some

302 individuals who received the flu shot does not prove that the flu shot caused the detection.[4] If the label

303 "Incorrect" warns us that there is something problematic with a piece of news, the label *post hoc* takes us

304 several levels deeper by allowing us to identify the level (reasoning) and the origin of misinformation.

305 Fallacies also take us in a new direction when observing and understanding broader trends in misinformation.

306 The taxonomy of ten fallacies that we employ, which is based on Tindale's (2007) framework, falls under four

307 broader classes: (a) fallacies from diversion, that divert the attention from the real issue at hand; (b) structural,

308 linked to the quantity of arguments; (c) logical; and (d) language fallacies. This broader categorization enables

309 us to understand patterns in the spread of misinformation. For example, there seems to be a correlation between

310 news based on the use of images and videos, and fallacies from diversion, especially "Red Herring" (the

---

[3] One of the earliest acknowledgements of cherry-picking appears in Plato *Hippias Minor* 369bc.

[4] This may be picked up in the detailed explanation provided by the fact-checkers but is not reflected in their labelling system. See https://healthfeedback.org/claimreview/claim-that-flu-shot-causes-false-positive-results-on-Covid-19-tests-is-unsupported-and-misleading/.

311 arguments are not relevant for the conclusion) and "Strawman" (when the other side's arguments are
312 intentionally misrepresented). In such cases, images or videos are taken out of their original context and are
313 employed as evidence for unrelated stories. An instance of such misinformation is re-labelling images of crowds
314 in demonstration as evidence for people rising up for Covid-19 related issues, whereas in fact those images are
315 taken out of their original context which has nothing to do with Covid-19 related demonstrations. Going beyond
316 the analysis and deconstruction of specific cases of misinformation, fallacy classes allow us to identify and
317 understand patterns in the spread of misinformation that can be peculiar to specific media and types of news.

318 **Data**

319 Our data comprises all the Covid-19 news that have been fact checked by the five fact-checkers in English:
320 *Snopes.com*; *Healthfeedback.org*; *Politifact.com*; *Fullfact.org*; *TheFerret.scot*. Our timeframe for the data
321 collection is from the beginning of the outbreak in January 2020 till end of June 2020, where we collected 1135
322 news articles. We have webcrawled the fact-checkers' official sites and created a dataset that contains the
323 following information: fact-checked news claim, link to the full fact-checked news, fact-checkers' comments,
324 and fact-checker' ratings.

325 It is important to emphasize that the various fact-checkers have different truth barometers in terms of number
326 and categories of ratings: while, for example, *TheFerret.scot*. uses a scale of seven ratings pointing to different
327 degrees of veridicality (e.g. "Mostly true", "false"), *Snopes* adopts a list of 14 ratings ranging from "Mixture" to
328 "Misattribution" or "Scam". Despite such variations, all the ratings allow to disentangle information deemed as
329 reliable (true information) information which constitutes complete fakery (disinformation) and information
330 which contains elements both of truth and of falsity (misinformation). Since *Fullfact.org* does not have a set of
331 fixed ratings tagged onto the fact-checked news, each fact-checked news article has been manually analyzed:

332 **Table 1**. Dataset of Covid-19 fact-checked news

Factchecker	Disinfo news	Info news	Misinfo news	TOT Fact checked News
Snopes	94	46	80	220
Health Feedback	2	0	68	70
The Ferret	27	0	13	40
Full Fact	46	31	208	285
Politifact	335	28	157	520
TOT	504	105	526	1135

334 In our dataset, disinformation constitutes 44% of the fact-checked news, true information amounts to 9% and
335 misinformation covers 46% of the cases, conforming that misleading news form a consistent portion of news
336 flagged as fake.

337 The review of the descriptions of the fact-checking processes disclosed by the fact-checkers reveals that there
338 are no common procedures for identifying which news to fact-check. However, we did identify several common
339 factors which influence the decision to choose news articles. These include: (a) newsworthiness; (b) popularity

340 across media; and (c) potential harm. As a result, our dataset of fact-checked news is not balanced as to topics

341 (e.g. symptoms vs governmental measures), but covers a wide range of domains.

342

343 **Methodology**

344 *Multilevel Analysis*

345 There is a proliferation of fallacy inventories associated with the various informal logic and rhetorical traditions

346 (Hansen, 1996). This diversity has so far hampered systematic annotation of fallacies. Aristotle, for example, in

347 his *Sophistical Refutations* (165b24-168a17), distinguishes fallacies dependent on the use of language and

348 expression (*in dictione*), such as the fallacies of equivocation and ambiguity, from those not dependent on

349 language (*extra dictione*), such as the fallacy of false cause. Pragmatic frameworks classify fallacies as

350 infringements of the rules of an ideal critical discussion (Van Eemeren et al., 2004). Regardless of the chosen

351 approach, the main issue at stake is the so-called *Fallacy Fork* (Boudry et al., 2015): cut-and-dry *compendia* of

352 fallacies are unlikely to be found in real life discourse. To cope with this, we have adopted a bottom-up

353 approach, with a focus on the analysis of the news articles in order to extract higher order insights. In this case,

354 the expert annotator analyzed 40 fact-checked articles randomly picked from climatefeedback.org, a platform

355 that gathers a network of scientists engaged in sorting fact from fiction in climate change media coverage, and

356 identified which fallacies have been called out through the comments of the reviewers. We intentionally focused

357 on news related to a topic detached from Covid-19 but of public interest to check whether the resulting

358 taxonomy is domain dependent or not.

359 As a starting point for our taxonomy of fallacies we adopted Tindale's (2007) framework, which gathers the

360 most common fallacies discussed in the informal logic tradition. The resulting annotation schema includes 10

361 types of fallacies scattered into four main groups: fallacies related to the presence of (sufficient) arguments:

362 *evading the burden of proof (EBP)*; fallacies pointing to the (un)intentional diversion of the attention from the

363 issue at hand: *strawman (ST)*, *false authority (FAUT)*, *red herring (RH)*, and *cherry picking (CP)*; fallacies

364 depending on the type of reasoning at play: *false analogy (FA)*, *hasty generalization (HG)*, *post hoc (PH)*, and

365 *false cause (FC)*; fallacies related to the language used: *vagueness (VAG)*. The guidelines contain the

366 description of the notion of fallacy and its relation to misinformation. Each fallacy is then defined, associated to

367 an example, and accompanied by one or more critical questions, which have turned out to be useful means to

368 evaluate arguments (Song et al., 2014). To offer a systematic and concise procedure, fallacies have been ordered

369 starting from those having to do with the quantity of information provided, followed by those related to aspects

370 external to the issue discussed; logical fallacies come into place after the other two classes are excluded. It is, in

371 fact, not worth looking at the type of reasoning at play if the information conveyed in the arguments is not

372 sufficient or irrelevant for the conclusion. The vagueness/ambiguity fallacy occupies the last position in the

373 heuristics when all the other options have been considered. In this way, the annotator can go through the critical

374 questions in a dyadic way, stopping when one of the critical questions is at stake:

375 Example (Kilpatrick & Fefferman, 2020)

376 Claim: "The WHO stated that asymptomatic spread of Covid-19 is 'very rare', therefore physical distancing and

377 face masks are not necessary"

378   Fact-checker comment:

379   "Imprecise: The scientific definition of the word "asymptomatic" refers only to a very small subset of infected

380   people who never develop symptoms during the course of their infection. However, the public tends to interpret

381   the word as also including presymptomatic individuals—those who are infected and not yet showing symptoms,

382   but eventually go on to do so. The WHO official was not referring to presymptomatic individuals in her

383   statement."

384

385   1. Does the news express an unassailable fact? Yes ---> ("REAL"NEWS); no ---->

386   2. Are there any evidence/arguments apart from the author's personal guarantee? Yes ---> 3; no --->

387   evading the burden of proof

388   3. Is the reported evidence (if any) the only available? Yes ---> 4; no --->

389   4. Is there any other data available which would bring to a different news? Yes ---> 5; no ---> cherry

390   picking

391   5. Are the evidence/arguments relevant for the news? Yes ---> 6; no ---> Red Herring

392   6. Is the news criticizing/rebutting somebody else's opinion? Yes ---> 7; No ---> 8

393   7. Is the criticized/rebutted opinion misrepresented? Yes ---> straw man; No ---> 8

394   8. Does the news contain an appeal to authority (e.g. scientist, politician etc.)? Yes ---> 9; No ---> 10

395   9. Did the authority make the attributed claim? Yes ---> 10; No ---> False Authority

396   10. Is the authority a genuine and impartial source? Yes ---> 11; No ---> False Authority

397   11. Does the news contain the comparison between two different situations? Yes ---> 14; No ---> 13

398   12. Are the two situations alike for real? Yes ---> 13; No ---> False Analogy

399   13. Are the similarities/dissimilarities relevant to prove the truth of the news? Yes ---> 14; No --->

400   False Analogy

401   14. Is the news a generalization drawn from a sample? Yes ---> 15; No ---> 17

402   15. Is the sample representative of the population? Yes ---> 16; No ---> Hasty Generalization

403   16. Is the considered sample relevant to the circumstances of a present situation or does it constitute an

404   exception? Yes ---> 7; No ---> Hasty Generalization

405   17. Does the news express a causal relation (cause/effect) between situations? Yes ---> 18; No --->

406   END("REAL"NEWS)

407   18. Is it possible that the situations co-occur by coincidence? Yes ---> POST HOC; No ---> 19

408   19. Could the situations be effect from separate or a common cause? Yes ---> FALSE CAUSE; No ---

409   > 20

410   20. Do concepts/words/phrases used in the news have multiple/vague/ambiguous meanings? Yes --->

411   Language fallacy; No ---> END("REAL"NEWS)

412

413   *Annotation of types of claims and types of sources*

414   Semantic types of claims have been analyzed to identify features that make a standpoint persuasive or predict

415   the types of arguments that are suitable to support them (Hidey et al., 2017). To investigate whether certain

416   types of claims circulated through news are more or less likely to convey dis-misinformation and/or to be

417 supported by fallacious arguments we have annotated the fact-checked news headlines using the following four

418 main categories:

- **Descriptions (D):** the claim expresses a factual state of affairs, i.e. ("There are x number of infections in London"; "the Oxford University lab has already produced a vaccine")

- **Predictions (P)**: The claim expresses a future state of affairs, i.e. ("The economy will end up being destroyed")

- **Interpretation (I)**: the claim expresses an explanation of states of affairs, i.e. ("The only reason why Italy has more cases, it is because they tested more")

- **Evaluation**: the claim expresses a more or less positive or negative judgement. Drawing from Liu (2012), evaluations are further classified as:

  - **evaluation-rational (ER)**: the claim expresses an opinion based on rational reasoning, non-subjective evidence or credible sources, i.e. "His phase 2 program is very solid"

  - **evaluation-emotional (EE)**: the claim expresses an opinion based on emotional reasons and/or subjective beliefs, i.e. "I don't' like having to use a mask at all times".

431 Our final layer of analysis consists in the annotation of the type of media source hosting the fact-checked news.

432 Due to the inherent fluidity of the digital medium, taxonomies cannot rely merely on medium factors observed

433 in computer mediated communication studies (Herring 2007): with the rapid evolution of technological

434 affordances features such as communication channels, synchronicity or message format are blurred. We have,

435 thus, decided to draw upon social and situational factors in defining our types of media sources. More

436 specifically, we have distinguished sources on the basis of the s*ocial practices*, "patterned ways of using

437 technologies and shared knowledge systems" (Yates & Sumner, 1997) and *discourse communities*, groups of

438 "reflexive actors with shared social practices and shared understandings of text types/genres, social contexts and

439 communicative acts", they give voice to: social media (e.g. Facebook, TikTok), broadcast media, blogs,

440 scientific articles, governmental sources (e.g. Liverpool City Council website). Among broadcast media we have

441 further distinguished broadcast media available through multi channels (e.g. Liverpool Echo), from those

442 available digitally exclusively since they potentially reach out to different audiences.

443 Our multi-level analysis has been carried out by two undergraduate students with no previous background in

444 Argumentation Theory or Informal Logic. They were introduced to fallacy theory and semantic types of claims

445 as well as the task guidelines through a 90 minute training session. They were given the same set of news in

446 CSV files and asked to identify: (a) type of semantic claim expressed in the headline; (b) type of source (e.g.

447 social media); and (c) type of fallacies (if any) at stake. The set of fact-checked news they assessed had been

448 rated between completely "true" (signaling information) and completely "false" (signaling disinformation).

449 They were also warned that a piece of news may contain more than one fallacy and asked to choose the one that

450 is more clearly flagged by the fact checkers. Once the annotators completed the annotation process, we asked a

451 rhetoric research specialist to go through the cases where the annotators disagreed and decide what label to

452 retain (this produced the *golden standard annotation* set). The guidelines and annotated files will be made

453 publicly available through the university data archive.

454 **Results of the analysis**

455 ***Results of the annotation*** *(RQ1)*

In order to evaluate the reliability of the annotations we have first calculated the inter-annotator agreement (IAA) using Cohen's κ (Cohen, 1960) since we have two annotators. To interpret the kappa values, we have relied upon Landis and Koch's scale, obtaining the following values:

Table 2. Inter-annotator agreement metrics

Level of analysis	*kappa* value	*Type of agreement*
type of media source	0.68	substantial
type of semantic claim	0.43	moderate
type of fallacy	0.52	moderate

The results show that while the types of media sources are easy to identify, the borders between types of claims and the types of fallacies are more blurred. This is not surprising since the *kappa* values are comparable with those obtained in tasks of similar complexity such as the annotation of argument schemes (Musi et al., 2016). It has to be remarked that our datasets constitues one of few annotated for fallacy type (Jin et al., 2022) Besides assessing the overall difficulty encountered by non-experts in using these analytical categories and offer a reliably annotated dataset, the main goal of the annotation was to understand what types of claims and what fallacies tend to be confused. On the one hand, different understandings of the semantics of news claims might trigger different decision-making processes: a piece of advice drawn from a claim perceived as descriptive is, for example, reasonably felt more reliable than one taken from a news expressing an interpretation. On the other hand, fallacies that are more challenging to identify are more likely to convey misinformation that is not recognized by the general public. To investigate these trends, we have built and analyzed the confusion matrices displayed in Tables 3 and 4.

Table 3. Confusion matrix for semantic type of claims

	D	P	I	EE	ER	TOT
D	679	26	9	10	12	736
P	106	65	3	14	0	188
I	73	6	33	13	2	127
EE	5	1	0	40	0	46
ER	22	0	1	2	13	38
TOT	885	98	46	79	27	1135

477

478

479

Table 4. Confusion matrix for fallacy types

	EBP	ST	FAUT	RH	CP	FA	HG	PH	FC	VAG	NO	TOT
EBP	73	2	4	3	2	0	1	0	0	1	13	99
ST	1	14	2	0	1	0	0	1	0	5	2	26
FAUT	2	1	11	2	2	0	2	1	0	2	5	28
RH	7	4	0	15	2	0	1	1	0	3	5	38
CP	8	2	3	3	39	0	3	1	1	2	16	78
FA	1	4	2	2	1	7	0	1	2	3	2	25
HG	15	5	4	1	10	1	47	1	2	6	7	99
PH	3	2	2	0	2	0	0	9	1	1	2	22
FC	0	0	0	1	1	0	1	0	2	1	0	6
VAG	7	1	3	5	11	1	6	1	0	51	13	99
NO	1	0	0	0	0	0	3	0	0	0	2	6
TOT	118	35	31	32	71	9	64	16	8	75	67	526

480

481 The analysis of the confusion matrix in Table 3 in comparison with the golden standard annotation revealed that

482 the category "description" has been overgeneralized by one annotator, covering cases where the claim expressed

483 instead a prediction, an interpretation, or an evaluation of the rational type. Zooming into those instances, it

484 seems that the cases that have been confused present as recurrent features a modal verb (e.g. "Fish tank additive

485 may treat coronavirus", prediction confused with description) or reference to an authority (e.g. "Italy is hit hard,

486 experts say, only because they have the oldest population in Europe" - interpretation confused with description;

487 e.g. The UK government no longer considers Covid-19 to be a "high consequence infectious disease" -

488 evaluation rational confused with description). This suggests that statements presented as possible states of

489 affairs that could, thus, happen in the future, have the potential to be misinterpreted as factual at the moment of

utterance; similarly, the ethos of authorities may lead to consider interpretations and evaluations as unassailable realities.

As far as fallacies are concerned, divergences between annotators are scattered across the full range, making it difficult to discern which pairs of fallacies tend to be confused more than others. However, it is clear that one of the two annotators had more difficulties in identifying a fallacy, especially when cherry picking is at stake. This is not surprising since the identification of arbitrary selection of sources requires a high degree of domain knowledge that is frequently hard to pinpoint, especially when available evidence has changed over time. It is, for instance, the case of the claim "Health authorities like the World Health Organization and the U.S. Centers for Disease Control and Prevention discourage people from wearing face masks" which expresses, as pointed out by *Healthfeedback.org* an outdated as well as partial recommendation since "Health authorities initially discouraged the public from wearing face masks due to extreme shortages of surgical and N95 masks needed to protect healthcare workers. However, health authorities now recommend mask use by the public, as new evidence suggests that cloth face masks worn by the public effectively reduce Covid-19 transmission." (Teoh, 2020).

The misinformation ecosystem (RQ2,3)

Solving cases of disagreement, the golden standard annotation has shed light on the misinformation ecosystem in our dataset of 1135 news articles. As to the types of sources, social media represent the large majority (72%) and feature multi-modal content ranging from tweets to YouTube videos. This is in line with research showing the privileged role of social media as vehicles of fake news (Mahid, Manickam & Karuppayah, 2018). Broadcast media cover 19% of the news with a preference for multi-channels news (12%) available, for instance, on digital as well as paper versions of the New York Times. Blogs represent the 6% of the sources encompassing personal as well as group pages. Finally, governmental sources constitute 3% of the sources including both national and regional official venues.

Turning to the semantic type of claims, descriptions cover three quarter (68%) of the cases, either presenting conspiracy theories as factual (e.g. "The Covid-19 coronavirus disease is spreading quickly from gas pumps.") or advancing misleading information about a wide variety of topics (e.g. "Eating bananas is a preventative against the Covid-19 coronavirus disease"). Claims of the interpretative type (14%) tend to express in our dataset causal relations where negative state of affairs related to Covid-19 are presented as effects of other supposedly co-occurring state of affairs; the cause-effect relation is for the most directly marked through a causal connective or phrase (e.g. "The (Covid-19) cases are going up, but it's **because** the testing is going up."; "96.3% of the Italy's Covid-19 deaths were actually caused by other diseases"). Regardless the form of expression, this configuration confirms the need for humans to engage in abductive reasoning (i.e., most probable conclusion based incomplete information) when fronting uncertain scenarios, looking out for what they consider best possible explanations for situations otherwise difficult to understand. Predictions (9%), expressed with higher epistemic commitment, have mostly scope over future directions taken by the pandemic (e.g. "Covid-19 is here to stay" and "we need to accept that and be prepared to deal with Covid long term") or outcomes of Covid-19 related policies (e.g. "The government in Oklahoma is planning to detain people unless they can show proof of vaccination"). Finally, among evaluative statements (9% overall), emotional evaluations (e.g. "While California is dying … Gavin (Newsom) is vacationing in Stevensville, MT!") outnumber (7%)

529   rational ones (e.g. "We've tested more than every country combined") confirming that appeal to fearmongering

530   is a common rhetorical strategy facilitating disinformation and misinformation spread.

531   When it comes to fallacies, the distribution across the 526 misinformation claims tagged as misinformation is

532   visualized in Figure 1:

Figure 1. Distribution of fallacies in our dataset

534
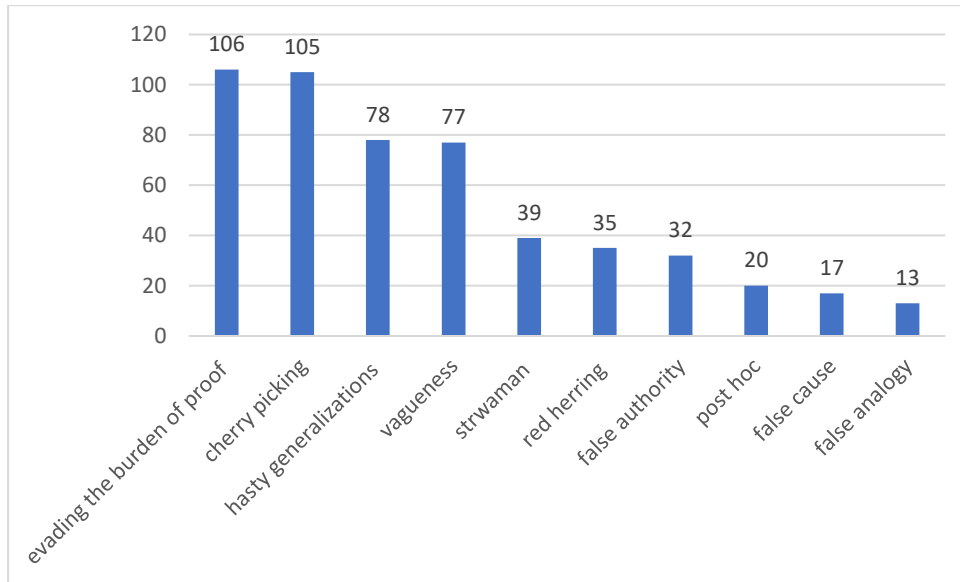
535   The lack of sufficient arguments in support of a claim (Evading the burden of proof) constitutes together with

536   the 'cherry picking' of evidence the most common fallacy in our sample, followed by generalizations drawn

537   from a non-representative or balanced sample (hasty generalization) and the use of vague/polysemous language

538   which allows for multiple interpretations (Vagueness). Arguments which misrepresent a third party's opinion

539   (Strawman) or appeal to an inappropriate authority are also quite frequent together with arguments that are

540   actually not relevant for the claim they support (Red Herring). Less common are the logical fallacies of Post

541   Hoc, where a correlation is presented as a causation; False Cause, where the wrong cause is attributed to an

542   effect and False Analogy, where a conclusion is drawn of the basis of similarities between two states of affairs

543   which are not comparable.

544   It has to be noted that the total number of cases containing fallacies amounts to 522 instead of 526 as in the

545   original annotation. This is because during the golden standard annotation process the expert annotator noticed

546   that certain instances have been considered by the two annotators as instances of misinformation, while

547   reporting no factual information to be classified as disinformation. Such cases stem from *FullFact,* that does not

548   include a fixed set of verdicts, and from cases labelled "Incorrect" in *Healthfeedback.org* (instead of False as in

549   other fact-checkers' truth barometers), as a further confirmation that the lack of a uniform set of verdict

550   descriptors hinders the recognition of different types of fake news.

551   Even though the restricted size of our sample prevents us from drawing any correlation between the frequency

552   of certain fallacy and the domain of the pandemic, it still suggests that the proposed taxonomy of fallacies bears

553   descriptive power when it comes to the grey area of misinformation under Covid-19 since for each news rated as

554   misinformation a fallacy has been identified by the annotators.

555   *Analysis of inter-level correlations (RQ4)*

556   To investigate the backbones of the misinformation ecosystem, we analyzed the mutual distributions of our

557   analytic categories throughout the dataset taking the golden standard annotation as a benchmark. Starting from

558   the semantic level, we obtained a positive statistical correspondence ($\chi2$ (36, n = 514) = 70.813, p = .0.000, with

559   a medium effect size Cramer'sV = 0.186) between the fallacy at stake and the type of claims that constitute the

560   main headline of the news. Looking at the residuals and contributions with highest value (Figure 2), three main

561   patterns stand out, namely interpretations * false cause; evaluation emotional * false analogy and prediction *

562   evading the burden of proof (RQ5):

563   **Figure 2.** Types of claim per fallacy – proportions and chi2 residuals and contributions

564
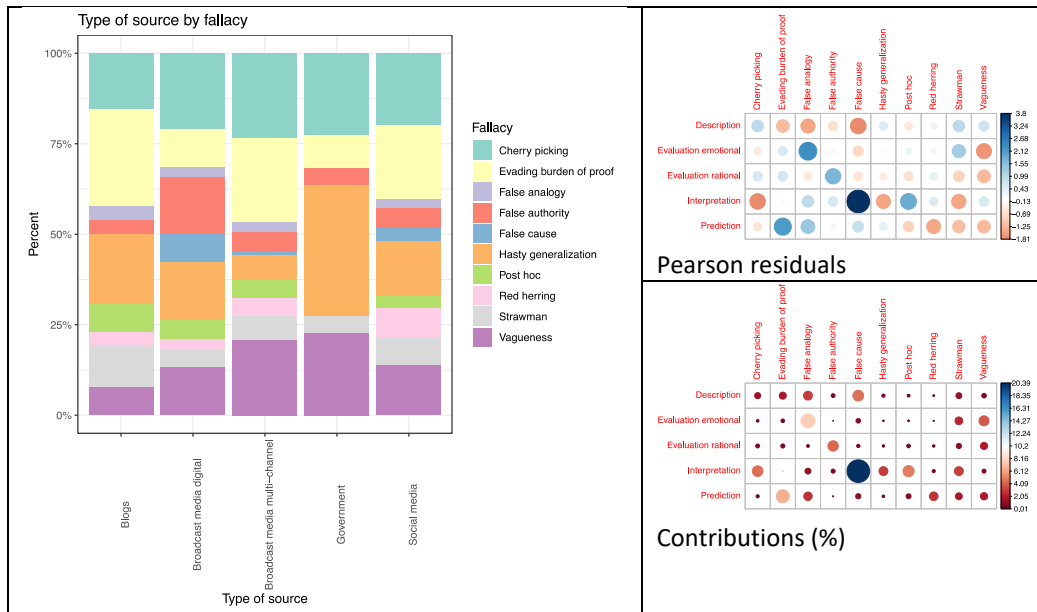
565   While it is expected that flawed causal relations would be used as arguments for faulty interpretations (e.g.

566   "There's a spike in [Covid-19] cases because there's a spike in testing", Facebook post) and that illegitimate

567   comparisons would fire up evaluative statements with a subjective connotation, the association between

568   predictions and evading the burden of proof is not intuitive. A predicament over a future state of affairs calls by

569   default for evidence to be credible. Closer examination of these cases reveals that such predictions relate for the

570   most to the decline of the virus with the warmer weather, drawing credence from people's hopes rather than

571   facts.

572   We found no statistical correspondence between the type of claims and the type of source: $\chi2$ (36, n = 514) =

573   22.544, p = 0.127. Though it should be noted that descriptive claims dominate all sources and that government

574   sources do not include evaluative claims (see Figure 3).

575

576

577

578

579

580

581

582

583

584

585

586

587

588   **Figure 3.** Types of claim per fallacy-percentages and residuals

589

Type of source by type of claim


590
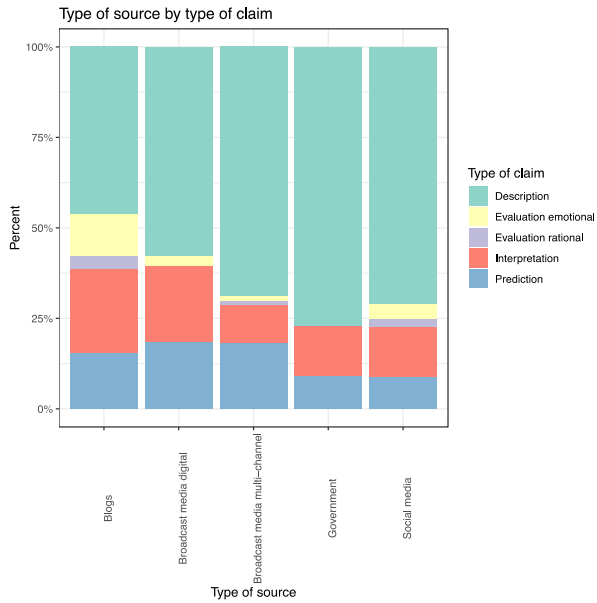
591   Blogs appear to have more subjective types of claims (evaluation emotional and interpretations). Comparing just

592   on subjective against non-subjective claims we find at statistically significant result at the p < 0.1 level (χ2 (4, n

593   = 514) = 8.116, p = 0.087, with a small effects size, Cramer's V = 0.126). In this analysis (see Figure 4) blogs

594   are the major contributor to the correspondence between factors. These results are limited by the nature of our

595   sample. We speculate that blogs are more evaluative sources in line with their nature as digital spaces working

596   as personal records. It is possible predictions tend to be preferred by broadcast media-multi-channels as the

597   focus is on future impacts. Further analysis of a larger sample of cases will be needed to assess any consistent

598   correspondence of sources and types of claim.

599   **Figure 4.** Types of claim per source-percentages and residuals
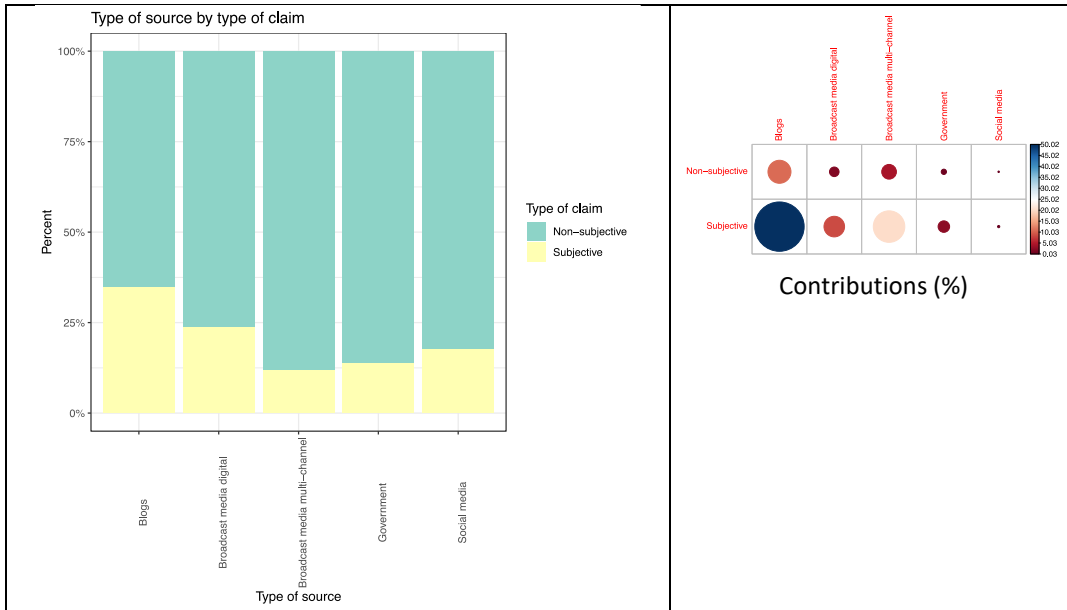
The other variable that corresponds significantly with our classification of types of sources is the broad category of misinformation, disinformation and information $(\chi 2\ (8, n = 514) = 33.139, p < 0.000, \text{small to medium effects size, Cramer's V} = 0.121)$. More specifically, while all the source types in our sample convey fake news as well as real news, social media and blogs constitute privileged channels for the spread of disinformation, while broadcast media and governmental official sources seem to be negatively correlated with blatantly false news. However, the trend is reversed when it comes to misinformation that bears positive residuals in correspondence with both broadcast media and government official sources:

Figure 5. Types of source across the misinformation ecosystem

609
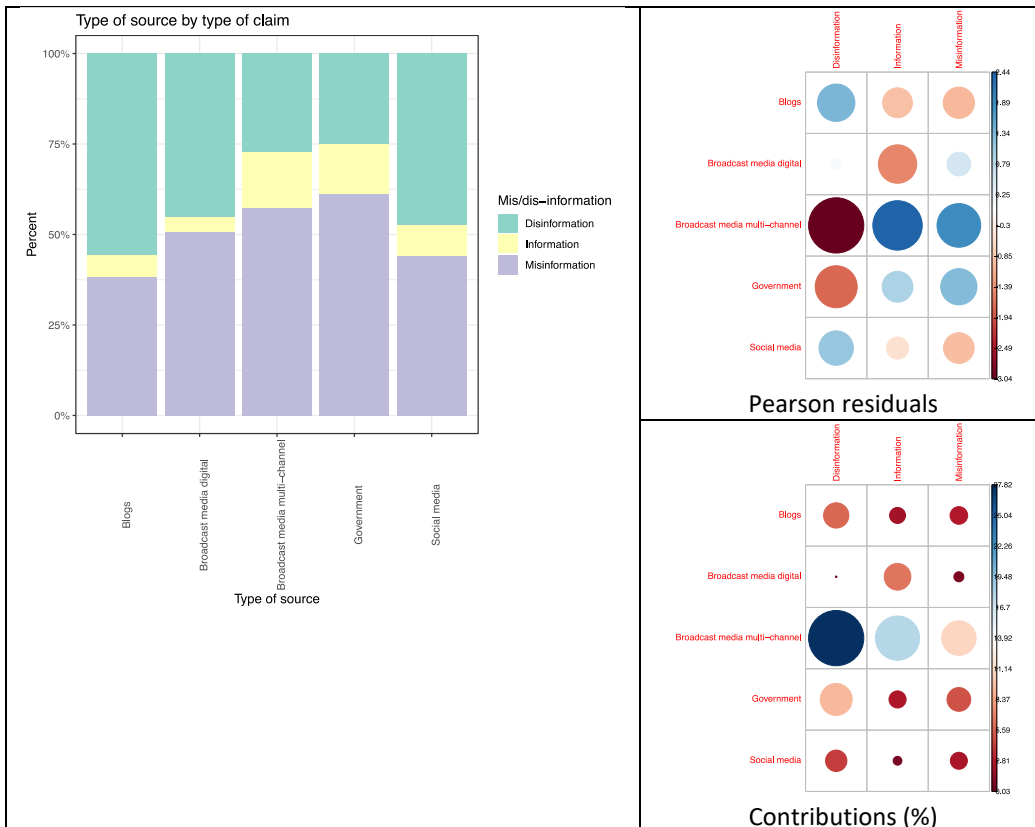
610   This trend partially aligns with results of studies showing that social media work as privileged vectors for the
611   spread of conspiracy theories/completely false information (Allington et al., 2020; H.O.Y. et al., 2020) and that
612   governmental communications spread confusing information which might cause misinformation (Kyriakidou et
613   al., 2020). The fact that misleading information can be spread by authoritative sources which are relegated to
614   gatekeeping processes reveals a gap between intentions and outcomes in radically uncertain situations such as
615   the pandemic. Looking at the distribution of fallacies (see Figure 6), we did not find any statistically significant
616   correlation between fallacy classes and types of sources ($\chi 2$ (12, n = 514) = 16.032, p = 0.190).

617
618
619
620
621
622
623
624
625   **Figure 6.** Fallacy class per type of source
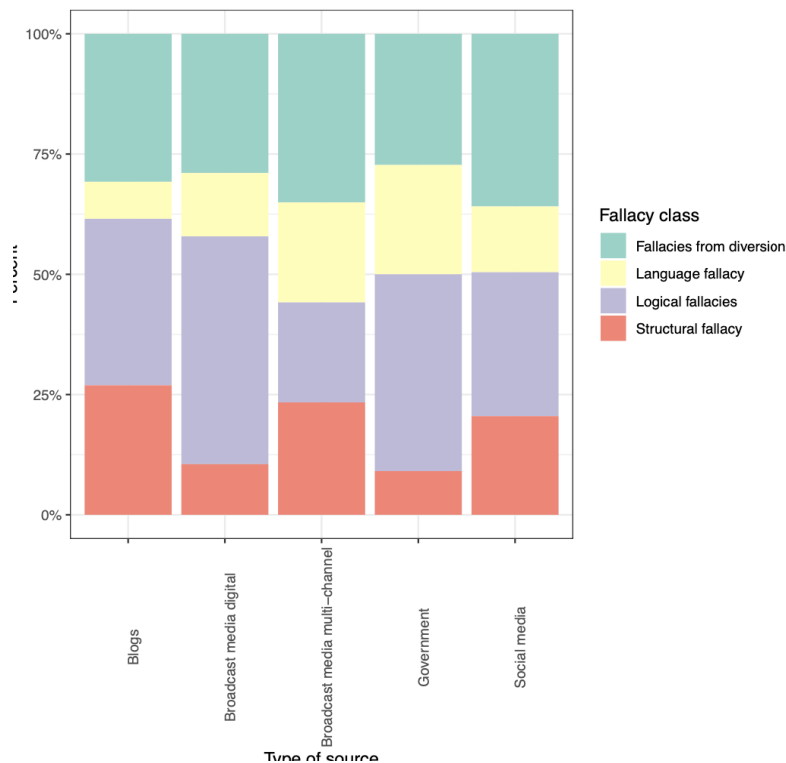
626
627   This result suggests that in crisis situations where epistemological differences between various publics (e.g.
628   journalists, policy makers, citizens) happen to be conflated, the entire range of fallacious moves is potentially
629   relevant across the board, regardless of the source. Official news media are in fact not exempt from the same
630   type of fallacious arguments spread by social media and blogs. From a methodological perspective this trend
631   also suggests that, even though qualitative categories such as that of fallacies allow us to operate a
632   categorization of the misinformation behaviors across media sources, it is not possible to calculate "averages"

633 and thus build reliable predictions without taking into account a variety of factors which go beyond single

634 variables. From the qualitative analysis of our sample, it has, for example, emerged that a factor influencing the

635 type of fallacious move at stake is the topic of the news: the strawman fallacy is mostly associated with news

636 about policies rather than symptoms or cures for Covid-19. Whether a policy-related statement is a good

637 candidate to become viral on social media or not, however, implies another set of factors which are hard to

638 predict.

639 **Conclusions**

640 In this study we address the phenomenon of fake news during the pandemic focusing on misinformation with

641 the aim of contributing to its systematic identification. Fact-checking misinformation, that is, information which

642 is misleading without necessarily containing false information communicated with the intention to deceive,

643 imposes even more challenges than identifying disinformation. On the one side, automatic fact-checkers are

644 currently unable to pick up information which may be factual, but misleading due to the lack of suitable training

645 data; on the other hand, human fact-checkers struggle to keep up with the proliferation of information across

646 digital media lacking a common truth barometer to flag the roots of misinformation. Drawing from the

647 awareness that fact-checking is not always a matter of facts, but frequently a matter of how arguments

648 supporting a news claim are built, we propose a discourse informed methodology to analyze misinformation

649 leveraging critical thinking and, more specifically, Fallacy Theory.

650 The underlying theoretical starting point is that fallacies, defined as arguments that seem valid but are not, work

651 as indicators of misinformation and provide more systematic explanations compared to mere labels as to why

652 news might be misleading. To verify the explanatory potential of fallacies and investigate the Covid-19

653 misinformation ecosystem, we adopt a bottom-up approach through the corpus analysis of a dataset of 1135 web

654 scraped fact-checked news in English and make the dataset publicly available to the scientific community. A

655 preliminary classification of the news according to the ratings shows that misinformation is more frequent than

656 disinformation across the fact-checked dataset.

657 We combine the annotation of fallacies, offering a novel heuristic procedure for their identification, with the

658 annotation of type of sources and semantic type of news claims. While we obtain successful inter-annotator

659 agreement metrics, the analysis of confusion matrices shows a tendency to overgeneralize the interpretation of

660 news claims as descriptions even when a prediction, an interpretation or an evaluation is at stake, especially in

661 the presence of a modal verb or a statement uttered by an authority (RQ1). Such results suggest that news

662 headlines have to be more clearly framed to disentangle opinions from reported facts.

663 As to the fallacies, cherry picking seems to be the most difficult to identify and not surprisingly so since it

664 requires a high level of epistemic vigilance and domain knowledge. The result of the golden standard annotation

665 allows us to come up with a decalogue of fallacies which exhausts our misinformation dataset pointing to flows

666 in the quantity and quality of arguments, the reasoning types at stake and the language used (RQ2). Besides

667 working as indicators of misinformation that could be used as features to build systems for the automatic

668 identification of misinformation, fallacies reveal the roots of misleading claims, being, thus, more informative

669 than truth barometers proposed by current human fact-checking enterprises. In this way, understanding fallacies

670 in social and broadcast media content may help people improve their digital literacy by learning how to cope

671 with such online manipulations in the future.

672 The inter-level analysis between types of sources, claims and fallacies reveals that there are significant
673 correlations between certain types of claims and fallacies as well as sources and that while social media are
674 privileged sources for disinformation, misinformation is spread across the board, calling for more careful
675 editorial processes in news production (RQ3, 4). The attested patterns offer guidance to sharpen critical thinking
676 when reading news, suggesting the need to keep epistemic vigilance high even when the sources are reliable
677 news media outlets and to ask ourselves questions when reading the news pointing, for instance, to the presence
678 of a sufficient number of arguments as well as the presence of correct inferences which do not, e.g., confuse
679 correlations with causations.

680 Interestingly, different types of fallacies do not pattern significantly with different types of sources showing that
681 in crisis situations such as the pandemic where certainty is not an option constitute a challenging information
682 environment for any kind of media. In such a post-truth scenario, audiences' digital literacy through critical
683 thinking offers a very important response to counter the infodemic. We believe that our decalogue of fallacies
684 constitutes a useful means to exercise audience's critical thinking towards reaching fake news immunity.

685
686 # REFERENCES

687 Abdulai, A. F., Tiffere, A. H., Adam, F., & Kabanunye, M. M. (2020). COVID-19 information-related digital
688 literacy among online health consumers in a low-income country. *International journal of medical
689 informatics*. *145*, 104322. https://doi.org/10.1016/j.ijmedinf.2020.104322

690 Allington, D., Duffy, B., Wessely, S., Dhavan, N., & Rubin, J. (2020). Health-protective
691 behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency.
692 Psychological medicine, 1-7. https://doi.org/10.1017/s0033291721000593

693 Amazeen, M. A., & Bucy, E. P. (2019). Conferring resistance to digital disinformation: The inoculating
694 influence of procedural news knowledge. Journal of Broadcasting & Electronic Media, 63(3), 415-432.
695 https://doi.org/10.1080/08838151.2019.1653101

696 Aufderheide, P. (1993). *Media Literacy. A Report of the National Leadership Conference on
697 Media Literacy*. Aspen Institute, Communications and Society Program, 1755 Massachusetts Avenue, NW,
698 Suite 501, Washington, DC 20036.

699 Beaunoyer, E., Dupéré, S., & Guitton, M. J. (2020). COVID-19 and digital inequalities:
700 Reciprocal impacts and mitigation strategies. *Computers in human behavior*. *111*, 106424.
701 https://doi.org/10.1016/j.chb.2020.106424

702 Boudry, M., Paglieri, F., & Pigliucci, M. (2015). The fake, the flimsy, and the fallacious:
703 Demarcating arguments in real life. *Argumentation*. *29*(4), 431-456. https://doi.org/10.1007/s10503-015-
704 9359-1

705 Brennen, J. S., Simon, F., Howard, P. N., and Nielsen, R. K. (2020). Types, sources, and claims of COVID-19
706 misinformation. *Reuters Institute*, *7*, 3-1. Available at: http://www.primaonline.it/wp-
707 content/uploads/2020/04/COVID-19_reuters.pdf.

708 Carmi, E. and Yates, S.J. and Lockley, E. and Pawluczuk, A. (2020). Data citizenship: rethinking data literacy in
709 the age of disinformation, misinformation, and malinformation. Internet Policy Review,[online] 9(2).

710   Available at: https://policyreview.info/articles/analysis/data-citizenship-rethinking-data-literacy-age-
711   disinformation-misinformation-and

712   Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological Measurement*.
713   *20*(1), 37-46. https://doi.org/10.1177/001316446002000104

714   Fletcher, R., Kalogeropoulos, A., Simon, F. M., and Nielsen, R. K. (2020). Information inequality in the UK
715   coronavirus communications crisis. Available at: https://reutersinstitute.politics.ox.ac.uk/information-
716   inequality-uk-coronavirus-communications-crisis.

717   Hansen, H. V. (1996). Aristotle, Whately, and the taxonomy of fallacies, in: *International Conference on*
718   *Formal and Applied Practical Reasoning*, Springer, Berlin, Heidelberg, pp. 318-330.
719   https://doi.org/10.1007/3-540-61313-7_82

720   Herring, S. C. (2007). A faceted classification scheme for computer-mediated discourse. *Language@ internet*,
721   *4*(1).

722   Hidey, C., Musi, E., Hwang, A., Muresan, S., & McKeown, K. (2017, September). Analyzing the semantic
723   types of claims and premises in an online persuasive forum. In *Proceedings of the 4th Workshop on*
724   *Argument Mining* (pp. 11-21).

725   Hobbs, R. (2010). *Digital and Media Literacy: A Plan of Action. A White Paper on the Digital and Media*
726   *Literacy Recommendations of the Knight Commission on the Information Needs of Communities in a*
727   *Democracy*. Aspen Institute. 1 Dupont Circle NW Suite 700, Washington, DC 20036.

728   Jin, Z., Lalwani, A., Vaidhya, T., Shen, X., Ding, Y., Lyu, Z., ... & Schölkopf, B. (2022). Logical Fallacy
729   Detection. *arXiv preprint arXiv:2202.13758.*

730   Kasprek, A. (2020, December 10). *No, mRNA COVID-19 Vaccines Do Not 'Alter Your DNA'* [fact-check of the
731   news claim *mRNA vaccines are capable of altering or damaging human DNA*.]
732   https://www.snopes.com/fact-check/mrna-alter-dna/.

733   Kilpatrick, A. & Fefferman, N. (2020, June 8). *People who do not show COVID-19 symptoms can and do*
734   *transmit it to others; physical distancing and face masks effectively reduce the risk of transmission.* [fact-
735   check of claims on Facebook *The WHO stated that asymptomatic spread of COVID-19 is "very rare",*
736   *therefore physical distancing and face masks are not necessary*]
737   https://healthfeedback.org/claimreview/people-who-do-not-show-covid-19-symptoms-can-and-do-transmit-
738   it-to-others-physical-distancing-and-face-masks-effectively-reduce-the-risk-of-transmission/

739   Li, H. O. Y., Bailey, A., Huynh, D., & Chan, J. (2020). YouTube as a source of information on COVID-19: a
740   pandemic of misinformation?. *BMJ global health*, *5*(5), e002604. https://doi.org/10.1136/bmjgh-2020-
741   002604

742   Mahid, Z. I., Manickam, S., & Karuppayah, S. (2018, October). Fake news on social media: brief review on
743   detection techniques. In *2018 Fourth International Conference on Advances in Computing, Communication*
744   *& Automation (ICACCA)* (pp. 1-5). IEEE. https://doi.org/10.1109/icaccaf.2018.8776689

745   Musi, E., Ghosh, D., & Muresan, S. (2016, August). Towards feasible guidelines for the annotation of argument
746   schemes. In *Proceedings of the third workshop on argument mining (ArgMining2016)* (pp. 82-93).

747   Kahne, J., & Bowyer, B. (2017). Educating for democracy in a partisan age: Confronting the challenges of
748   motivated reasoning and misinformation. *American Educational Research Journal*, *54*(1), 3-34.
749   https://doi.org/10.3102/0002831216679817

Koltay, T. (2011). The media and the literacies: Media literacy, information literacy, digital literacy. *Media, culture & society*, *33*(2), 211-221. https://doi.org/10.1177/0163443710393382

Kyriakidou, M., Morani, M., Soo, N., & Cushion, S. (2020). Government and media misinformation about COVID-19 is confusing the public. *LSE COVID-19 Blog*.

Yates, S., Carmi, E., Pawluczuk, A., Lockley, E., Wessels, B., & Gangneux, J. (2020). Me and My Big Data Report 2020: Understanding citizens' data literacies: thinking, doing & participating with our data.

Yates, S. J., & Sumner, T. R. (1997, January). Digital genres and the new burden of fixity. In *Proceedings of the Thirtieth Hawaii International Conference on System Sciences* (Vol. 6, pp. 3-12). IEEE.

Ramsetty, A., & Adams, C. (2020). Impact of the digital divide in the age of COVID-19. *Journal of the American Medical Informatics Association*, *27*(7), 1147-1148. https://doi.org/10.1093/jamia/ocaa078

Song, Y., Heilman, M., Klebanov, B. B., & Deane, P. (2014, June). Applying argumentation schemes for essay scoring. In *Proceedings of the First Workshop on Argumentation Mining* (pp. 69-78).

Teoh, F. (2020, December 19). *The U.S. National Childhood Vaccine Injury Act does not stop people from suing vaccine manufacturers* [fact-check of the webpage https://archive.ph/OYp8S] https://healthfeedback.org/claimreview/the-u-s-national-childhood-vaccine-injury-act-does-not-stop-people-from-suing-vaccine-manufacturers/

Teoh, F. (2020, August 4). *Health authorities encourage the general public to wear face masks to reduce COVID-19 transmission* [fact-check of claims on Facebook *Health authorities like the World Health Organization and the U.S. Centers for Disease Control and Prevention discourage people from wearing face masks*] https://healthfeedback.org/claimreview/health-authorities-encourage-the-general-public-to-wear-face-masks-to-reduce-Covid-19-transmission/

Tindale, C. W. (2007). *Fallacies and argument appraisal*. Cambridge University Press.

Van Eemeren, F., Grootendorst, R., & van Eemeren, F. H. (2004). *A systematic theory of argumentation: The pragma-dialectical approach*, Cambridge University Press.

Vraga, E. K., Bode, L., & Tully, M. (2020). Creating news literacy messages to enhance expert corrections of misinformation on Twitter. *Communication Research*, 0093650219898094. https://doi.org/10.1177/0093650219898094