Computers in Human Behavior Reports 2 (2020) 100043

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

journal homepage: www.journals.elsevier.com/computers-in-human-behavior-reports

Tailoring heuristics and timing AI interventions for supporting news veracity assessments

Benjamin D. Horne [a], Dorit Nevo [b,*], Sibel Adali [c], Lydia Manikonda [b], Clare Arrington [c]

[a] School of Information Sciences, University of Tennessee, United States
[b] Lally School of Management, Rensselaer Polytechnic Institute, United States
[c] Computer Science, Rensselaer Polytechnic Institute, United States

ARTICLE INFO

Keywords:
Fake news
Misinformation
Heuristics
AI

ABSTRACT

The detection of false and misleading news has become a top priority to researchers and practitioners. Despite the large number of efforts in this area, many questions remain unanswered about the ideal design of interventions, so that they effectively inform news consumers. In this work, we seek to fill part of this gap by exploring two important elements of tools' design: the timing of news veracity interventions and the format of the presented interventions. Specifically, in two sequential studies, using data collected from news consumers through Amazon Mechanical Turk (AMT), we study whether there are differences in their ability to correctly identify fake news under two conditions: when the intervention targets novel news situations and when the intervention is tailored to specific heuristics. We find that in novel news situations users are more receptive to the advice of the AI, and further, under this condition tailored advice is more effective than generic one. We link our findings to prior literature on confirmation bias and we provide insights for news providers and AI tool designers to help mitigate the negative consequences of misinformation.

1. Introduction

Fake News consists of fabricated, misleading information that is intended to deceive (Jang & Kim, 2018). Fake news is increasingly prevalent in online platforms and in particular on social media (Allcott & Gentzkow, 2017). To detect fake news, there has been significant work on developing AI methods with the aim of supporting news consumers' assessments of the veracity of news articles. However, there is relatively little work on whether such methods are indeed useful in end-user facing systems. Such success, even if the algorithmic advice is highly accurate, is not straightforward.

The acceptance of algorithmic advice might depend on the individual's prior beliefs about the news topic, how well-established those beliefs are, how the advice is provided, and what cues are received from significant others (Colliander, 2019; Moravec et al., 2018). The acceptance of such advice also depends on the level of uncertainty and risk associated with the news topic. It has been shown that such uncertainty can lead to an information overload, as well as the increased prevalence of rumors, conspiracy theories, and disinformation (Starbird et al., 2020). A case in point is the outbreak of COVID-19, which was classified as an international public health emergency in January 2020 by the World Health Organization. Soon after this announcement, the WHO described information on COVID-19 as an *'infodemic'* due to "an over-abundance of information - some accurate and some not".[1] Due to the high frequency of new, evolving, and sometimes conflicting information being reported during such emerging situations, news consumers are tasked with making sense of changing streams of news across multiple technology mediated environments.

Stepping back from the COVID-19 crisis and focusing on efforts to mitigate misinformation in general, these typically fall into three broad categories: 1) reducing visibility of unreliable information, 2) educating information consumers to better evaluate misleading information, and 3) flagging unreliable information when it is shown to the user, potentially with warning labels and corrections. In this paper we focus on the latter, and we explore how effective warning labels can be designed.

Past research has shown that simple fake news flags, without any specific presentation or explanation are not always effective (Horne et al., 2019b; Moravec et al., 2018). Therefore, we build on the use of *heuristics*

* Corresponding author.
E-mail addresses: bhorne6@utk.edu (B.D. Horne), nevod@rpi.edu (D. Nevo), sibel@cs.rpi.edu (S. Adali), manikl@rpi.edu (L. Manikonda), arrinj@rpi.edu (C. Arrington).

[1] https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200415-sitrep-86-covid-19.pdf?sfvrsn=c615ea20_6.

https://doi.org/10.1016/j.chbr.2020.100043
Received 6 September 2020; Accepted 15 November 2020
Available online 8 December 2020

(or mental shortcuts) that can help users make better judgments and consequently can reduce the reach and virality of misinformation. For example, using a simple heuristic such as distrusting information because others distrust it was shown to affect the reaction to- and the spread of-fake news, more so than a generic disclaimer did (Colliander, 2019). Similarly, research shows that when people are prompted to think about specific heuristics they can better evaluate information (Bago et al., 2020; Moravec et al., 2020).

This paper delves deeper into the use of heuristics for news assessment and how they can be harnessed to provide better warning labels. Specifically, we explore the heuristics that are employed for news assessment and then offer targeted AI advice that focuses on these heuristics. In light of the literature that argues that more subjective heuristics, such as previously held beliefs and knowledge, may play a major role in news assessment (Broockman & Kalla, 2016), we add a novel contribution to understanding heuristics in news assessment by studying the impact of our targeted AI advice in two news settings. The first, which we term the *everyday news* situation, presents respondents with articles on vaccinations and climate change. The second, which we term the *emerging news* situation, presents respondents with COVID-19 news articles. We chose these two settings because we expect them to differ in the extent and strength of previously held beliefs, and the openness of news consumers to accept external advice.

In sum, we focus on the problem of understanding the effectiveness of AI advice in fake news situations under varying conditions. We test different forms of advice based on heuristics that are used by news consumers. We also test different news situations to understand how advice is accepted given previously held beliefs and knowledge.

This two-fold objective necessitates a sequential empirical assessment. Consequently, in a first, qualitative, study we focus on understanding the use of heuristics in the specific context of news assessment. Next, we use the insights obtained from this qualitative exploration to develop a more focused empirical study to test the impact of AI interventions under different conditions. Our findings show that users rely on a broad set of heuristics for assessing news, and that these heuristics typically confer to the general types identified in prior literature. In a second study we then show that indeed AI advice can lead to better judgement of news articles as legitimate or fake, but only under the emerging news situation, in which readers don't already have strongly held beliefs. Further, we find that tailored, rather than generic, AI advice is more effective in impacting readers' opinions.

These results offer two important contributions to the literature. First, we demonstrate the difference in veracity intervention effectiveness across emerging and non-emerging news situations. Our work demonstrates that interventions in news consumption may be less effective when the news contains recurring topics, where strong prior beliefs may already be formed. However, in novel and evolving topics, such as crisis events, interventions can be effective. Second, our results show that utilizing already established mental shortcuts and reasonings (heuristics) in veracity interventions can be more effective than using generic veracity flags, an important finding for practitioners.

In what follows we provide the broad literature foundations for our work, focusing on the use of heuristics in general, and for news assessment in particular. We also explain how AI assessment of news content is typically carried forth and outline our proposed approach of incorporating more tailored heuristics into AI advice. Next, we describe two studies of news assessment. In the first study, we employ qualitative analysis of news consumers' justification of their assessments and we elicit emerging codes pertaining to the heuristics that used by those readers. We then use those insights in a more focused theoretical development to offer two hypotheses, which we then test in a second study. In this study, we explore how advice that is tailored at specific heuristics impacts news judgement in two different news situations. We conclude with a discussion of our insights, contributions, and implications for future research.

2. Background

The increasing availability of news in online media has led to a true cultural change in the consumption and production of news. News sources have shifted to an all-day news production schedule to meet changes in consumer demand (Boczkowski, 2011). Changes in revenue streams have led to the closure of many local newspapers and the reduction of investigative journalism budgets. Many alternative news outlets have begun low-cost, online publishing operations, which often do not conform to the same journalistic principles and quality control as well-established print media. Motives across these alternative information sources range from profit to political motives, both of which are less concerned with long-term credibility and lack an underlying commitment to truth. This complex news and media landscape has inevitably led to increasing availability of misleading and outright false information.

Traditionally, news producers have relied on journalistic standards of accuracy, objectivity, transparency, and accountability to produce credible information and foster trust in the readers.[2] But these standards don't always exist in current times. Often, sources will act in coordination, copying information from each other and employing many other techniques to push certain narratives to prominence (Horne et al., 2019d; Starbird et al., 2018). Additionally, social media forums have become a location where a significant number of individuals get their news, either through shared news stories or by shared claims that exist outside of news articles. The further interaction between viral sharing patterns and social engagement based recommendation algorithms has made processing and assessing information credibility challenging.

These are the challenges that news consumers face on a daily basis. Dealing with those challenges requires a significant amount of cognitive effort that may not always be available to them. Consequently, individuals often turn to simple rules of thumb, or heuristics, when assessing news, or they turn to seek help through technology, and in particular through algorithmic advice and flagging of fake news. We review both of these approaches next.

2.1. A technological approach

Recently, many researchers and developers have focused on building AI tools to detect or approximate the veracity of news articles, with the aim of providing assistance to news consumers in detecting false and misleading news. In these directions, the predominant focus is on developing classifiers of fake and non-fake news and improving their accuracy. The majority of models developed have utilized signals of veracity found in the text of news articles, where these signals are extracted using various techniques in Natural Language Processing (Baly et al., 2018; Cruz et al., 2019, pp. 999–1003; Hosseinimotlagh & Papalexakis, 2018; Bozarth & Budak, 2020). Furthermore, these models have ranged widely in the machine learning algorithms used, from decision tree methods to deep learning methods. An example of one such tool is NELA (Horne et al., 2018, 2019a, p. 2019a), which is a news veracity classifier that utilizes text features from six categories: style, complexity, bias, affect, moral, and event. These feature groups range from lexicon-based features to more complex language processing features. Another example tool is DeClarE (Popat et al., 2018), which utilizes automatic text feature extraction using a bidirectional LSTM, as well as combining both article-level and claim-level models.

Regardless of the feature extraction method used, when training algorithms to detect fake news, training data must be labeled as fake or legitimate. This requires the researcher to either pass their own judgment on the veracity of news or to rely on outside journalistic organizations for labeling. Commonly, researchers use labeling that is done at source-level using 3rd party journalistic organizations, such as NewsGuard or Media

[2] https://www.americanpressinstitute.org/journalism-essentials/what-is-journalism/elements-journalism/.

Bias/Fact Check (MBFC) (Gruppi et al., 2020; Nørregard et al., 2019). For example, NewsGuard has journalists rate news sources based on a nine weighted criteria: (1) The source does not repeatedly publish false content, (2) The source gathers and presents information responsibly, (3) The source regularly corrects or clarifies errors, (4) The source handles the difference between news and opinion responsibly, (5) The source avoids deceptive headlines, (6) The source's website discloses ownership and financing, (7) The source clearly labels advertising, (8) The source reveals who is in charge, including any possible conflicts of interest, and (9) The source provides information about content creators. In this ground truth rating system, the weights of the criteria add up to 100 if a news source passes all of them. This scale from 0 to 100 can then be divided into two news source veracity categories. NewsGuard divides the sources above 50 as credible and below 50 as not credible, but this threshold can be adjusted or sources near the threshold border can be left out of training to make a stricter veracity model. MBFC uses a similar process, but provides more categories of news sources, such as news sources with specific political bias or news sources that often push pseudoscience. There has also been attempts to learn from more granular ground truth, such as individually fact-checked claims or articles. These labels also typically come from 3rd party journalistic organizations, such as Snopes or PolitiFact (Hassan et al., 2017; Shu et al., 2020; Wang, 2017). While there has been some success with models using claim-level or article-level ground truth, there are concerns of accuracy of these models over time (Horne et al., 2019a) and their access to labeled data for emerging events.

In all of these works, the goals have been computational in nature, focusing on increasing the accuracy and robustness of detection methods. What is missing from the existing literature is an assessment of the extent to which these tools impact news consumers' perception of the veracity of the news that they read. That is, while the accuracy and performance of the classifiers is commonly studied, their ability to affect news consumers has not been significantly explored. Of the existing literature on AI's effectiveness in human decision making, very few studies have explicitly examined advising news consumers. These studies (e.g. Horne et al., 2019b; Moravec et al., 2018) have experimented with generic AI advice, such as presenting flags or labels for fake news articles or providing generic AI predictions of veracity for news articles, and have shown mixed results. For example, while presenting a flag on fake news increased the time consumers spent considering the veracity of the news, it did not have a significant impact on the judgment of veracity (Moravec et al., 2018). When AI advice was given as a generic, probabilistic statement (i.e. "AI System says this article has an 8% chance of being reliable"), there were significant impacts on the judgment of veracity, but those impacts varied widely across individual differences, such as expertise and social media use (Horne et al., 2019b).

Negative outcomes are also possible. For example, it has been shown that attaching warnings regarding correctness to some articles may cause readers to assume articles without any attached warnings are accurate (Pennycook et al., 2020). Furthermore, attempting to correct strongly held beliefs may lead to strengthening of these beliefs (the so-called backfire effect), although this effect has not been consistently reproduced (Swire-Thompson et al., 2020).

To address the above limitations, we propose to look at how individuals assess news using simple rules of thumb and how these can be incorporated back into the algorithmic advice to make it more effective. We review the concept of heuristics next.

2.2. A heuristic approach to news judgement

The term "heuristic" stems from Greek (εὑρίσκω) which means, "to find". In cognitive psychology it is used as a useful shortcut, or a rule of thumb for searching through possible solutions (Hoffrage & Reimer, 2004). More generally, a heuristic is defined as a "rule of thumb for making a decision, forming a judgement, or solving a problem without the application of an algorithm or an exhaustive comparison of all available options" (Colman, 2015). Heuristics are mental shortcuts that ease the cognitive load on decision-makers (Myers & DeWall, 2018).

Using heuristics for making a judgement is not guaranteed to be optimal or rational, but is sufficient to achieve an immediate goal when finding an optimal solution is impossible. Simon (1956) coined the term "satisficing" to describe a situation where people seek solutions or accept judgements that are "good enough" for their purposes, but that could be optimized. Simon also showed that humans operate within "bounded rationality" where decision-makers' rationality is limited due to different constraints. Regardless of the solution being not guaranteed optimal, heuristics are extensively used in decision making and have been studied in various contexts.

Kahneman et al. (1982) reveal how when thinking under uncertainty, biases can reveal some heuristics, such as (i) *representativeness*, which is usually employed when people are asked to judge the probability that an object or event A belongs to class or process B; (ii) *availability* of instances or scenarios, which is often employed when people are asked to assess the frequency of a class or the plausibility of a particular development; and (iii) *adjustment from an anchor*, which is usually employed in numerical prediction when a relevant value is available. These heuristics are highly economical and usually effective, but they lead to systematic and predictable errors. A better understanding of these heuristics and of the biases to which they lead could improve judgments and decisions in situations of uncertainty.

Gigerenzer and colleagues studied broad types of heuristics, such as *fast and frugal* heuristics (Gigerenzer & Goldstein, 1996), which gained attention to describe human judgement; *recognition* heuristic (Goldstein & Gigerenzer, 2002) and *take-the-best* heuristic (Gigerenzer & Goldstein, 1999), which have since been modified and applied to domains from medicine, artificial intelligence, and political forecasting (Czerlinski et al., 1999; Graefe & Armstrong, 2012). Such heuristics are recognized as having ecological validity for being effective and rational choices for decision making under certain scenarios (Todd & Gigerenzer, 2007), especially those that involve reasoning under uncertainty (Lieder et al., 2018; MacGillivray, 2017). For example, in the medical domain, these heuristics can help doctors and patients make better decisions by concentration on few relevant predictors for a given condition (Marewski & Gigerenzer, 2012).

Typically, heuristics are not used in the same way across the board, with factors affecting the use of heuristics including age (Besedeš et al., 2012), autonomy (Blumenthal-Barby, 2016), and specific contexts. For example, different heuristics would be used in political decision making (Lau & Redlawsk, 2001), judicial decisions (Fischhoff et al., 2002; Rachlinski, 2000), parenting (Davidson, 1995; Renjilian et al., 2013), medicine (Bodemer, 2015), weather forecasting (Doswell, 2004), and credibility of the news (Horne & Adali, 2017; Rubin et al., 2016; Tandoc Jr et al., 2018).

In this study we are interested in understanding decision heuristics in the specific context of news consumption. Narrowing in on the relevant literature, Table 1 summarizes key articles within the news consumption context, the heuristics they have explored, and the decision-making context they studied.

The above table shows that heuristics are commonly used in both news selection and news evaluation decisions.

Table 1 shows that the majority of studies on news selection have explored social heuristics, looking for cues from significant others. Indeed, research shows that when searching for information, individuals explicitly seek and trust social contacts who hold similar beliefs in online media. These contacts provide social confirmation of their existing beliefs (Metzger et al., 2010). This selective search and evaluation of information and curation of social contacts can create echo chambers that significantly reduce the diversity of viewpoints that users are exposed to in social media platforms (Vicario et al., 2016).

In news evaluation, the more prevalent heuristics are self (cognitive) heuristics, as well as attributes of the news content itself and the identity of the source. Self, or cognitive, heuristics focus on the need for cognition

Table 1
Heuristics in the context of News Consumption.

Authors	Heuristics explored	Heuristic type	Decision-making context
Anspach (2017)	Different levels of social media activity that is attributed to different sources ranging from fictional individuals to own friends and family members of the subjects were considered as features to investigate how social media affected individuals' news selection. Experiments suggest that *online endorsements* and *discussions* serve as heuristics when deciding the kind of content to consume.	Significant others	News selection.
Sundar et al. (2007)	Three distinct cues of news were studied – 1) name of the primary source of information, 2) time elapsed since the news story broke and 3) number of news articles written about this news story by other news organizations in the context of *machine* (if a mere machine chose the story) and *bandwagon* (if so many news organizations think this is news, then it must be") heuristics.	Content and source	News selection.
Knobloch-Westerwick (2005)	The paper argues that in the news selection process, people are less likely to follow the heuristics of *popularity indications* suggesting why the *bandwagon effect* may not apply. Through their experiments this work suggests that cognitive circumstances, situational perceptions, or inter-individual differences help guide certain individuals to let popularity indications guide their selections whereas this is not the case for others highlighting "follow the crowd" is not the only option.	Significant others	News selection.
Messing and Westwood (2014)	Social media provides readers a choice of stories coming from different sources recommended by politically heterogeneous individuals highlighting the aspect of social value. This study builds on existing models of news selectivity showing that the distinctive features of social media such as *social endorsements* trigger several *decision* heuristics and are more powerful heuristic cues compared to the source of information.	Significant others	News selection and evaluation.
Chung (2017)	Explored the heuristics of *media credibility* (H1) or *social metrics* (H2) when evaluating news online. H1 – those with low personal relevance will report higher news evaluations when reading a news story from a high credibility news organization; H2 – those with low personal relevance will report higher news evaluations when reading a news story with social media metrics.	Media	News evaluation.
Grabe et al. (2000)	By utilizing different packaging styles, emotional and physical arousal of viewers were tested along with encoding and retention of information by leveraging the *limited capacity model* of information processing.	Content	News evaluation.
Kim and Dennis (2019)	The paper examines the effect of presentation format and highlighting of *source identity* on believability of news.	Source	News evaluation
Metzger and Flanagin (2013)	The article focuses on the use of *cognitive* heuristics in evaluating the credibility of information in online environments compared to other heuristics such as *reputation*, *endorsement*, *consistency*, *self-confirmation*, *expectancy violation* and *persuasive intent*. *Cognitive* heuristics constitute information processing strategies that ignore information to make decisions more quickly and with less effort than more complex methods, and thus they reduce cognitive load during information processing.	Self	New credibility.
Go et al. (2014)	Three types of heuristic cues were utilized in this study to understand online	Self and significant others	Online news perception.

(*continued on next page*)

Table 1 (*continued*)

Authors	Heuristics explored	Heuristic type	Decision-making context
	news perception. They are: 2 *expertise* cues – low vs high, 2 *identity* cues – in-group vs out-group and 2 *bandwagon* cues – low vs high.		
Dvir-Gvirsman (2019)	The process of news selection where users when presented with different types of information directed towards persuading them to read a post was evaluated by using *social* cues as the process heuristic which are less demanding compared to text-based cues. Two traits have been studied in this context – *need for cognition* and *self-monitoring*.	Self	Persuasion to read news.
Igartua and Cheng (2009)	The paper conceptualizes the *framing effect* as a heuristic process to understand how peripheral cues in the news story guided information processing. This refers to a process involving these operations: selecting and emphasizing words, expressions and images, to lend a point of view, focus or angle to a piece of information.	Content	News framing.

and are typically based on previously held beliefs, knowledge, and world views. An important feature of such heuristics is that they selectively use information (ignoring what is deemed irrelevant) in order to make decisions more quickly and with less effort (Metzger & Flanagin, 2013). Finally, heuristics that are rooted in the content of articles or in the identity of the source focus on attributes such as source credibility, recency, or supporting evidence (e.g. Sundar et al., 2007). It has been shown that media literacy approaches that providing readers with a set of rules and content heuristics can increase readers' ability to differentiate between mainstream and false news (Guess et al., 2020).

2.3. Enhancing technology with heuristics – research question

There is increasing evidence that heuristics can be usefully invoked in news veracity judgments. Specifically, Kim et al. (2019) showed that displaying both aggregated expert ratings of articles at the source-level and aggregated non-expert ratings of articles at the source-level had strong effects on veracity judgments. Similarly, Kim and Dennis (2019) showed that nudging consumers to think about the source of a news article, yet another heuristic being provoked (i.e. Is the source of the information trustworthy?), made users more skeptical of news articles, regardless of the credibility of a source. More recently, Moravec et al. (2020) found that when fake news flags are designed to be processed by System 1 cognition (heuristics), they can be effective without training the consumer to be aware of the flag. With awareness training, the authors found both flags designed to trigger heuristics and those designed to trigger deliberate cognition were effective (Moravec et al., 2020).

These previous studies point to the potential usefulness of tailoring AI advice to already established decision making heuristics in fake news interventions. Hence, in this paper, we explicitly study this notion by introducing AI advice that is tailored to confirmed heuristics used by news consumers. Specifically, we address the following research question: How can heuristics be incorporated to increase the effectiveness of AI advice in news veracity assessment?

The answer to this question is not straightforward. It requires first a stronger understanding of specific heuristics that are employed in news veracity decisions, which can then serve as a foundation for developing specific testable hypotheses. To this end, we first conduct a qualitative exploration of news veracity heuristics and then build on our insights to develop a more focused research model, that we test in a second study.

3. Study 1: exploring news heuristics

When faced with a news veracity judgement task, individuals will use heuristics to support their decision. We reviewed two broad types of such heuristics in the section above, namely cognitive and content heuristics.[3] To validate the use of these heuristics in the specific context of our study, we begin with a qualitative investigation of the heuristics used by news consumers. The objective of this study is to gain a clearer understanding of specific news evaluation heuristics and set the stage for the experimental study to follow.

3.1. Design

The task presented to respondents in this study was to read one randomly selected news article from our data set. After reading the article, respondents were asked whether or not they believed this article and why.

We conducted the study on Amazon Mechanical Turk (AMT), a commonly used platform for data collection. Although AMT might not always be the most fitting platform, studies have shown that it is suitable for use in settings similar to ours (open ended subjective assessments), and generates quality data (Bates & Lanza, 2013). Given that cultural differences may impact our finding, especially given our news context, we limited responses to workers from the United States only. To ensure quality responses we also limited the HIT approval rate (a common quality measure on AMT) to 99%.

For this study we used ten different articles covering two everyday news topics – climate change and vaccinations – and we used articles that presented arguments both for and against these two topics. Three of the articles rejected the issue of climate change and they came from Natural News, Jew World Order, and the Gateway Pundit; two articles presented an anti-vaccination stance and they came from Freedom Bunker and Natural News; two articles supported the issue of climate change and they came from NPR and Fortune; finally, three articles presented a pro-vaccination stance and they came from BBC, Chicago-Sun Times, and NPR. The articles are shown in Table 4 under the description of Study 2 below.

In selecting articles of varying ground truth we used 3rd party organizations, similar to previous literature (Gruppi et al., 2020; Nørregard et al., 2019). Since our study has participants reading individual news articles, our ground truth must also be at the article-level. To this end, we utilized a three step process, in which we (1) found sources that were labeled as reliable and unreliable by Media Bias/Fact Check (MBFC), (2) found topic specific articles from those sources (climate change and vaccination), (3) selected articles that have been fact checked by a 3rd party journalistic organization, such as Snopes, PolitiFact, FactCheck.org, Washington Post Fact Check, or AP Fact Check.

[3] We exclude the social heuristics here as our focus is on news evaluation rather than selection.

Table 2
News heuristics.

Heuristic	Definition	Example
Personal belief alignment	Respondents mention an alignment (or lack thereof) between the views of the article and their personal beliefs	*I believe climate change is happening and we are running out of time to try to save our planet* *The article is anti vaccination which is an incorrect and dangerous sentiment.* *Vaccinations are proven safe and effective and everyone should get vaccinated.*
Personal experience alignment	Respondents mention an alignment (or lack thereof) between the views of the article and their past experiences (past experience is explicitly mentioned)	*I believe it because I use the oil and it works. There are studies that show it works.* *Because I had chickenpox when I was young*
Previous knowledge alignment	Respondents mention an alignment (or lack thereof) between the views of the article and their past experiences (prior knowledge is explicitly mentioned)	*It seems to correspond with other factual known information.* *Because I have heard a lot of this on other news sources previously.* *It contradicts things I know are facts*
Supporting evidence provided in article	Respondents mention that the article provides (in the body of the text) supporting evidence from external sources	*The article doesn't give any reputable sources and I am skeptical when I can't find a link from an actual medical site.* *The article provides detailed information that is verifiable from other sources.*
Bias perception	Respondents explicitly mention that the article seem biased (or unbiased).	*… Also the article is based on science and particularly the last paragraph where the author states that more research is needed makes me more inclined to believe the contents of the article because it shows a lack of bias. this information or news seems very biased against left wing members or news.* *It's seems very biased. Funny they mention "fake news and fake science" when they seem to base their claims on just that.*
Accuracy perception	Respondents explicitly mention that the article seem accurate (or inaccurate).	*It seems accurate.* *I believe it because it seems factual*
Coherent Story	Respondents mention that the article is written in a coherent and factual manner. The story is logical.	*I believe the information in the article because it presents logical arguments. Each person is important and changing ourselves is often the first and best way to enact any kind of change.* *Lays out claims and backs them up.* *It's incoherent. Sounds more like a rant.*
Writing Style	Respondents comment on the writing style and writing quality of the article from a grammar and language standpoint	*They use derogatory terms such as libtard which is an obvious sign that the article is biased.* *It is written poorly and very emotionally and full of hyperbole and it goes against all the established science. because it doesn't used slanted language it only presents facts.*
Trusted Source	Respondent explicitly mention their opinion about the source of the article	*NPR is a great source and I have always trusted content from them* *The Chicago-Sun Times is a reputable paper.* *I am not sure of how reliable the source is.*

The article length (number of words) ranged between 117 and 1686 words, with an average length of 781.33 words, and they were assigned to respondents at random.

As mentioned above, we asked respondents whether or not they believed the article, and why. The first question was answered using a five-point scale with options ranging from "definitely yes" to "definitely not". The second question was presented as an open text response box. We also collected information on the following controls: age, gender, level of education, news sources commonly used (social media, news websites, TV, newspaper, other), frequency of news consumption, typical social media used, news sharing frequency on social media, and other trusted news sources. Finally, we included one reliability question to ensure people were paying attention to the task.

3.2. Data collection

We piloted the questionnaire in November 2019, using an initial batch of 14 responses. We ensured that our payment of $0.5 was sufficient and that the questions were clear to the users. No changes were made following this pilot, and we proceeded to collect a first batch of 94 responses. We collected a second batch of 99 responses in January 2020. Finally, as the pandemic broke and the world around us changed we decided to collect a third batch of 87 responses in April 2020. We were concerned that people's news consumption and assessment might have changed during these times and we wanted to capture any such changes in our data.

Of the 294 total responses 22 were incomplete, leaving us with a final data set of 272 responses. The demographics of those who have not completed the survey were not significantly different than of those who completed it, eliminating concerns of non-response bias. Tables A1 and A2 in the Appendix summarize the demographic and control information collected from respondents in the four data collection rounds. What we learn from these two tables is that there was no significant difference in our respondents to the four rounds of data collection. There was also no significant difference in terms of news consumption habits on these two topics due to the worldwide pandemic.

Once all of the results were received, we compiled the list of arguments provided by respondents on the reason why they believed, or did not believe, the article they read. We then coded these responses in two steps. First, two of the authors individually went over the data to identify broad codes, and they continued until the list of codes was stable (no new codes emerged). Example for such codes are "trusted source" "bias" or "alignment with personal belief". The two authors then met and discussed their coding to develop the combined list of codes, which is shown in Table 2 in the Findings section. Next, the two original coders and one other author individually coded the full data based the identified heuristic codes that were developed. A participant response could be assigned one or more codes, per the judgment of each coder. We then measured initial agreement among the four judges, which was 81.4%. The judges then met to clarify definitions and discuss disagreement. Two of the categories "alignment with personal belief" and "alignment with prior knowledge" had lower agreement rate because they interfered with each other. Another conflict arose in classifying articles into the "accuracy" or "coherent story" heuristic. After discussion we clarified the definitions for each category and the three judges returned to the data for another coding round. At the end of this round, agreement by all three judges increased to 93.0%. We examined the items that still had mixed votes and found that they made too broad statements and did not identify clear heuristics. For example: "This just seems to be an ad trying to sell a product!", "I believe it is true.", and "It's an opinion article and not news." Therefore, with the agreement rate beyond 90% and the weak agreement items carefully reviewed, we proceeded to analyze the data.

3.3. Findings

Table 2 presents the heuristics we identified from the text, along with

Table 3
Heuristics prevalence.

Do you believe the information in this news article?	Belief alignment	Experience alignment	Knowledge alignment	Supporting evidence	Bias	Accuracy	Coherent story	Writing style	Trusted source
Total in percent	29%	3%	22%	11%	12%	6%	6%	11%	14%
Heuristic type	Self/Cognitive Heuristics			Content Heuristics					Source heuristic

Table 4
Articles used.

Everyday News		
Source	Title	Ground Truth
Natural News	Climate change HOAX has literally convinced a member of Congress that "the world is going to end in 12 years"	Not Credible
Freedom Bunker	Fight illness with this ancient immune booster	Not Credible
Jew World Order	Greenpeace Founder: Global Warming is a Hoax	Not Credible
	Pushed by Corrupt Scientists 'Hooked on Government Grants'	
The Gateway Pundit	NOAA Ruins Assertions by Unhinged Democrats that Global Warming Has Caused Increase in Hurricane Activity	Not Credible
Natural News	World Health Organization declares *anti*-vax movement to be a top "global health threat" just like the climate change hoax … the vaccine deep state grows desperate	Not Credible
BBC	'Completely avoidable' measles outbreak hits 25-year high in US	Credible
NPR	Climate Change Was the Engine That Powered Hurricane Maria's Devastating Rains	Credible
Chicago-Sun Times	Kentucky governor exposed his kids to chickenpox instead of getting vaccine	Credible
NPR	New U.S. Measles Cases Break 25-Year-Old Record, Health Officials Say	Credible
Fortune	U.S. Carbon Emissions Soared in 2018. Here's Why	Credible
Emerging News		
Source	**Title**	**Ground Truth**
The New York Times	Open Windows. Don't Share Food. Here's the Government's Coronavirus Advice.	Credible
Reuters	World Faces Chronic Shortage of Coronavirus Protective Equipment: WHO	Credible
The Guardian	Can a face mask stop coronavirus? Covid-19 facts checked	Credible
Breitbart	Hillary Clinton Falsely Claims to Jimmy Fallon That Trump Called Coronavirus Outbreak a 'Hoax'	Not Credible
Natural News	Vitamin C infusions being studied in China as possible treatment for coronavirus-related pneumonia	Not Credible
Natural News	Spirulina found to boost the body's type 1 interferon response to fight RNA viral infections "including coronavirus," new science finds	Not Credible
The Russophile	CORONAVIRUS HOAX: Fake Virus Pandemic Fabricated to Cover-Up Global Outbreak of 5G Syndrome	Not Credible
The Russophile	CORONAVIRUS SPECIAL REPORT: Worldwide Outbreaks of 5G Syndrome and 5G Flu Driving Pandemic	Not Credible
The Liberty Daily	Coronavirus: Chinese Espionage Behind Wuhan Bioweapon?	Not Credible

their definitions and example quotes. Note that we focused on whether the heuristic was mentioned and did not focus on only positive (e.g. "I trust this source") or only negative (e.g. "I do not trust this source") values. As long as "source" was mentioned, for example, we coded the argument as using the source heuristic. Next, we examine the prevalence of each code in our data set, as shown in Table 3. For example, 29% of respondents have indicated reliance on previously held belief in making their news veracity judgement, whereas only 11% where looking at the availability of supporting evidence within the article. Below the numbers in Table 3 we also indicate the type of heuristic, based on our literature review. Three of the heuristics used: previously held knowledge, beliefs, and experience are classified under the self/cognitive heuristic type. Five are rooted in the content of the article and are therefore classified as content heuristics. Finally, another important heuristic that emerged from our study is the source heuristic, which we maintained as a separate category.

4. Discussion

The tables below provide two important insights for the remainder of our work. First, we identify the set of heuristics that are employed by news consumers to decide whether or not they believe a given article. These heuristics are a mix of the self/cognitive heuristics (e.g. personal belief), content heuristics (e.g. bias and accuracy), and the source heuristic. Hence, the study validates the use of these heuristics in our specific context. Second, we note that the cognitive heuristics are more prominent in people's decisions. That is, news readers will default to believe articles that align with their previously held knowledge and beliefs. We also found that the identity of the source is an important heuristic as shown in previous literature (Pornpitakpan, 2004).

When we drilled deeper into the use of the source heuristic, we found that it was more prominently used in positive news judgement. That is, people tend to trust news from sources that they deem reliable. In negative judgements, people relied more heavily on other heuristics, specifically writing style, perceived bias, and (mis)alignment with prior beliefs. This relationship between the valence of the evaluation and specific heuristics can be investigated further in future research.

Focusing on the balance between the cognitive heuristics and the content and source heuristics, we now turn to the question of the effectiveness of algorithmic advice in fake news interventions and develop our specific hypotheses.

5. Hypothesis development

While the AI can provide direct signals to support the use of content and source heuristics, signals that address cognitive heuristics are often not effective because of confirmation bias. Confirmation bias means seeking or interpreting evidence in ways that agree with existing beliefs and expectations (Nickerson, 1998). Minas et al. (2014) show that individuals tend to disregard information that challenges their pre-existing views and pay greater attention to supporting information. Confirmation bias is the tendency to interpret new information such that information that supports pre-existing views is considered while information that challenges those views is ignored (Minas et al., 2014). This bias can be attributed to overconfidence of people in evaluating the correctness of their knowledge (Koriat et al., 1980).

Due to confirmation bias, beliefs that are based on cognitive heuristics can be difficult to change as individuals may overlook and undervalue information that refutes their beliefs (Metzger & Flanagin, 2013). This helps them avoid the discomfort caused by contradicting information, also referred to as the cognitive dissonance (Festinger, 1957). Furthermore, confirmation bias can weaken the effect of algorithmic advice. Moravec et al. (2018) for example, found that a fake news flag did not influence users' beliefs, and they attributed this to the existence of confirmation bias. In their words "the flag was not enough to overcome

participants' inherent confirmation bias" (p. 21).

Building on their above, we put forth two hypotheses that focus on how the AI can effectively provide advice by utilizing pre-existing source and content heuristics and avoiding the limiting effect of confirmation bias. We focus on two important aspects of the advice provision, which we term: *timing* and *tailoring*.

Timing refers to when the algorithmic advice is offered in terms of the novelty of the information. Specifically, in our news context timing refers to whether the news situation presented is familiar (or "everyday" news) or emerging. Prior research has shown that confirmation bias is strong when prior beliefs or knowledge are strong (Park et al., 2013) and when one has high confidence in their decision ability (Rollwage et al., 2020). Confirmation bias is reduced under information disfluency or difficulty in attaining and processing information (Hernandez & Preston, 2013). Further, since significant cognitive effort might be required to change prior beliefs that are already stored in memory, these might persist even after receiving the corrective information (Ecker et al., 2015). This is why supplying corrective information regarding news veracity at the time of exposure is particularly important (Swire & Ecker, 2018).

In emerging news situations, news consumers face extreme uncertainty concerning the evolving situation, accompanied by information overload and increased prevalence of rumors, conspiracy theories and disinformation (Starbird et al., 2020). This serves to reduce the confidence in decision ability and increase news disfluency. With confirmation bias reduced, we expect that the effectiveness of algorithmic advice to increase, as news consumers will rely more heavily on such advice:

H1. the effectiveness of algorithmic advice will be higher in emerging news situation than in everyday news situations

Tailoring refers to the presentation of the algorithmic advice in a specific way. For example, a generic fake news flag is considered untailored, because it does not provide any explanation to news consumers. We specifically focus on the explanation provided alongside the algorithmic advice for two reasons: 1. Prior studies have shown that AI advice with explanation performed better in shaping users' opinions than AI advice without explanation (Horne et al., 2019b) and 2. Prior studies have shown that formatting advice to present source information in different ways affected the believability of articles (Kim & Dennis, 2019; Kim et al., 2019). Indeed, explainable AI is a growing topic of interest in technology and human behavior research (e.g. Holzinger et al., 2017) and explanation approaches vary depending on the biases they try to eliminate (Wang et al., 2019). Examples of explanations provided can be prototype instances of specific decision outcomes, explanations of the prevalence of specific outcome, and providing the attributes that played a role in the decision outcome (Wang et al., 2019).

From a theory perspective, the dual processing theories such as Elaboration Likelihood Model, Heuristic Systematic Model, and Kahneman's "System 1 and System 2" model (Chaiken, 1999; Kahneman, 2011; Petty & Cacioppo, 1986) provide support for the use of heuristic based explanations in AI advice. All three theories argue that when lacking in motivation and ability to deeply process messages, readers will turn to use heuristics. News consumption, particularly on social media, often is not motivated consumption. More precisely, information on social media feeds is often consumed passively, where the consumer is not actively searching for news, but rather browsing a feed that is a mixture of various types of information including news (Aula. 2005; Boczkowski et al., 2017; Hertzum & Frøkjær, 1996). The infinite-scroll design or, in terms of Thaler and Sunstein's Nudge Theory (Thaler & Sunstein, 2009), the choice architecture of social media platforms encourages passive information consumption, hence news may be indirectly recalled later, information may be partially consumed, or only the titles of news articles are read (Horne et al., 2019c; Wang et al., 2016). Boczkowski et al. (2017) describe this concept clearly, saying: "young users consume news on social media can be characterized with the notion of "incidental news": most young users get the news on their mobile devices as part of their constant connection to media platforms; they encounter the news all the time, rather than looking for it; but click on them only sporadically and spend little time engaging with the content."

Given this passive nature of modern news consumption, the use of mental shortcuts in veracity belief and information recall is common. Using this idea that heuristics are often used in news veracity assessments, we hypothesize that:

H2. the effectiveness of algorithmic advice will be higher when the advice is tailored to specific heuristics as opposed to a generic advice

To test the above hypotheses we designed a second study varying the conditions of news novelty and advice presentation. We describe this study and its results next.

6. Study 2: tailored AI advice under two news conditions

With the enhanced understanding of the heuristics that news consumers use, in addition to the literature we reviewed, and in line with the above two hypotheses, we set up to study the effect of timing and tailoring on news readers' ability to assess the veracity of articles.

6.1. Design

6.1.1. Timing

We employed two different conditions to assess the effect of the timing of AI intervention. In the first condition, which we term Everyday News, we presented respondents with one of ten articles on climate change or vaccinations, assigned at random (the same articles that were used in Study 1). These articles are shown in Table 4 below. The second condition, which we term Emerging News, presented respondents with one of nine articles on COVID-19, assigned at random (also shown in Table 4). The selection of articles within each condition was described under study 1, with the same process used to select the additional set of COVID-19 news. The selection of specific articles was random, we just ensured that we have coverage of both sides of the topic and varying ground truth. The two conditions, Everyday and Emerging news, were guided by our hypothesis that the AI interventions' will be more effective in more novel situations.

6.1.2. Tailoring

The second hypothesis focuses on the effect of tailoring the AI advice rather than providing a generic one. This followed evidence from the literature that simple fake news flags, without any specific presentation or tailoring to heuristics, were not always effective (Horne et al., 2019b; Moravec et al., 2018). To test this hypothesis, we included four specific AI tailoring conditions:

1. *No AI*: only the articles were presented to respondents who were then asked to make their judgement (as we did in study 1).
2. *Generic AI*: at the top of the page we presented one of the following two statements: "Our smart AI system believes this article" or "Our smart AI system does not believe this article".
3. *AI Source*: at the top of the page we presented one of the following two statements: "Our smart AI system indicates this is a trusted news source" or "Our smart AI system indicates this is a not a trusted news source".
4. *AI Content*: at the top of the page we presented one of the following two statements: "Our smart AI system rates this article as accurate and reliable" or "Our smart AI system rates this article as inaccurate and unreliable"

The above conditions were designed based on our study of heuristic types, with conditions 3 and 4 touching on the two heuristics types of source and content, respectively, to increase the external validity of our findings.

After presenting one of the above statements we asked respondents to read the article and use the five points scale to indicate whether they

Table 5
Agreement coding example.

Source	Ground Truth	Title	Do you believe the information in this news article?	Coded as
The Liberty Daily	FALSE	Coronavirus: Chinese Espionage Behind Wuhan Bioweapon?	Definitely not	Agree
The Russophile	FALSE	CORONAVIRUS SPECIAL REPORT: Worldwide Outbreaks of 5G Syndrome and 5G Flu Driving Pandemic	Might or might not	Not coded
Natural News	FALSE	Vitamin C infusions being studied in China as possible treatment for coronavirus-related pneumonia	Definitely yes	Disagree
Reuters	TRUE	World Faces Chronic Shortage of Coronavirus Protective Equipment: WHO	Probably not	Disagree
The NYTimes	TRUE	Open Windows. Don't Share Food. Here's the Government's Coronavirus Advice.	Probably yes	Agree

believed the information in the article (the scale ranged from "definitely yes" to definitely not"). We then again provided an open text box for respondents to provide their reasoning for why they believed (or did not believe) the article.

6.2. Data and measures

We collected data for this study in two phases, both using AMT. We maintained the settings to include only respondents from the United States with a HIT approval rate of at least 99%. In phase 1, we collected data using our Everyday News set of articles. For Condition 1 (No AI) we used our responses from study 1, since it utilized the same set of articles. There were 272 responses in study 1 that referred to the ten articles used in this study. We then collected additional 83 useable responses for the Generic AI condition, 159 useable responses for the AI Source condition, and 137 useable responses for AI Content condition. The demographics of respondents in this study were similar to those of the respondents in study 1, as were the social media and news consumption characteristics. We present those in Tables A3 and A4 in the Appendix.

In Phase 2, we collected data for our Emerging News set of articles. We collected 114 useable responses for condition 1 (No AI), 113 useable responses for condition 2 (Generic AI), 104 useable responses for condition 3 (AI Source), and 115 useable responses for condition 4 (AI Content). Again, the demographics of respondents in this study were similar to all other responses (see Tables A3 and A4 in the Appendix).

To draw conclusions about the *effectiveness of algorithmic advice,* which is our dependent variable in the above two hypotheses, we computed the percentage of agreement between respondents' classification of the article and its ground truth, under the different AI conditions. Because respondents provided their judgement on articles using our standard five-point scale we first converted their responses to a simple True or False scale. Recall that the question we presented respondents was "Do you believe the information in this news article?". Responses of "definitely not" and "probably not" were converted into *false* and responses of "definitely yes" and "probably yes" were converted into *true.* The midpoint of our scale "might or might not" was left as is and was not used for this analysis.

We then compared these newly created true/false rating to our ground truth and marked whether it was in agreement or disagreement with it. For example, if a respondent said they "definitely not" believe an article which we also marked as False in our ground truth assessment, then we recorded an agreement. If a respondent said they "probably yes" believed an article that we marked as False, then we recorded a disagreement. We show examples of this coding in Table 5 below.

Finally, we computed the proportions of agreement and disagreement under each of the three conditions, and each of the two phases, and we conducted pairwise tests of proportions.

6.3. Findings

Table 6 shows the proportion of agreement under each condition, and the pairwise comparisons tests for both news settings conditions.

To understand Table 6 note, for example, that the 84% under p_1 in the first row means that 85% of respondents in the Everyday News and No AI condition correctly identified the ground truth of the article.

Table 6
Results of pairwise comparisons of proportions.

Everyday News					
Comparing		p_1	p_2	z stat	Sig.
No AI	Generic AI	84%	82%	0.50	0.31
No AI	AI Source	84%	83%	0.25	0.40
No AI	AI Content	84%	84%	0.11	0.46
Generic AI	AI Source	82%	83%	(0.28)	0.39
Generic AI	AI Content	82%	84%	(0.37)	0.36
AI Source	AI Content	83%	84%	(0.11)	0.45
Emerging News					
Comparing		**p_1**	**p_2**	**z stat**	**Sig.**
No AI	Generic AI	72%	84%	(2.04)	**0.02**
No AI	AI Source	72%	93%	(3.59)	**0.00**
No AI	AI Content	72%	93%	(3.65)	**0.00**
Generic AI	AI Source	84%	93%	(1.74)	**0.04**
Generic AI	AI Content	84%	93%	(1.79)	**0.04**
AI Source	AI Content	93%	93%	(0.04)	0.48

Hypothesis H1 stated that the effectiveness of algorithmic advice will be higher under the Emerging News condition than under the Everyday News condition. To test this hypothesis, we can look at the comparison between No AI and Generic AI under both parts of Table 6. As can be seen the effect is significant for the Emerging News condition ($p = 0.02$) but not for the Everyday News condition ($p = 0.31$), supporting H1.

Hypothesis H2 stated that the effectiveness of algorithmic advice will be higher when then advice is tailored as opposed to generic. To test this hypothesis, we can look at the difference between the agreement proportion between the Generic AI condition and either the AI Source or AI Content conditions. We can see that this effect is significant, but only under the Emerging News condition ($p = 0.04$ for both the AI Content and AI Source conditions in the Emerging News condition, $p = 0.39$ and 0.36 for the two respective comparisons in the Everyday News conditions). This partially supports H2.

Overall, our results show that there were significant differences in respondents' ability to correctly identify the ground truth of an article when provided the AI advice, but more importantly – when provided tailored AI advice on a specific heuristic, and only in the case of emerging news. The between group differences in the Everyday News condition were not significant, reflecting the fact that respondents were able make reasonably good judgments on their own without the advice of the AI (Condition 1). Further, even when given tailored AI advice these assessments did not significantly improve, remaining at 83% and 84% for both heuristics used. These results support the literature that confirmation bias might play a stronger role than any other heuristics, *when respondents hold strong prior beliefs and are confident in their knowledge.*

In the Emerging News condition, we found that respondents were less

able to make correct assessments on their own, with only 72% agreement with the article's ground truth[4]. When provided generic AI advice, respondents' ability to identify fake news (across all nine emerging news articles) improved significantly to 84%. It further improved when we tailored the AI advice to either the source heuristic (93%) and the content heuristic (93%). The willingness of respondents in this group to accept the AI advice is one indication of weaker confirmation bias, per our literature review.

Before we explore specific comments that were provided by respondents to justify their news assessment, we set to better understand the difference between the two conditions by examining the types of heuristics that were prevalent under both. To this end, we coded the open-ended comments for both data sets and reviewed the distribution of heuristics, as shown in Table 7. For this analysis we only used the No AI condition, which included respondents who did not receive any AI flag. This serves to prevent any bias that might be presented by our experimental conditions and the specific AI advice we provided.

Using a Chi-Square goodness of fit test on the distribution of heuristics in Table 7 we find that the distribution for Emerging News is significantly different than that of Everyday News ($\chi^2 = 33.94$, $p < 0.01$), with the difference stemming predominantly from the Bias, Accuracy, and Consistent Story heuristics. What we also see, however, is that there is no significant difference in the use of the cognitive heuristics between the two conditions, at least not in terms of their prevalence. Our next step is to examine individual comments to understand any differences in the *strength* of these heuristics. Before we do this, however, as a final analysis we examined the average number of heuristics used by each respondent under both conditions (recall that the justifications provided by respondents could be coded as employing one or more heuristics). In the Emerging News condition, the average number of heuristics used by respondents was 1.3, compared with 1.17 in the Everyday News condition, and this difference was significant using an upper tail *t*-test, at the 0.05 level. In other words, people used slightly more heuristics in the Emerging News condition than in the Everyday News condition. This can be further indication that specific heuristics were not as strong, in and of themselves, under this condition.

6.3.1. Evidence of confirmation bias

A qualitative investigation of comments demonstrates that confirmation bias played a likely strong role in the assessment of Everyday News articles. We first note that many of the respondents who relied on cognitive heuristics under this condition often relied *solely* on those. For example, a respondent who correctly did not believe an article which we flagged as False noted:

> *I don't believe that global warming is made up, the scientific consensus is very strong that it is happening and is caused by human behavior and carbon emissions. The fact that the author is calling into question this basic principle makes me question a lot of their other assertions, including the fact that Greenpeace has been "hijacked by the extreme left."*

This respondent is using his or her belief about global warming to form an opinion about the article. The strength of this belief is reflected in it being well informed and articulated and serving as a foundation for the overall article judgement. Similarly, a respondent who *did believe* an article which we flagged as False used a similar justification strategy in disagreeing with the AI advice:

> *Global warming and climate change are lies. These lies have been promoted since the 1800's. NONE of their predictions have panned out. NONE of their predictions will ever pan out. What happened to the ozone hole crisis? What about the melting ice caps? They've grown back to be thicker than ever in recorded history. How about those rising oceans that have been predicted over and over and over and over again? Shouldn't there be actual evidence of claims made since the 1800's?*

We can again see the strength of this belief through the use of capitalization, examples and questions, and the sole use of the cognitive heuristics in judging the article.

This type of justification was not unusual in the Everyday News condition. The use of cognitive heuristics and likely confirmation bias was apparent across all comments provided by participants regardless of the ground truth of the article (although respondents were quite adept at correctly identifying this ground truth). While other heuristics were also used in some comments, such as the reputation of the source or the writing style of the articles, confirmation bias was strongly evident in many of the comments. Examples for such mixed use of heuristics are provided below:

> ***Cognitive heuristic:*** *I believe the general thrust of the article is true that we are facing a climate emergency that will only become more severe in the coming years and some kind of action needs to be taken to ameliorate it ….* ***Content heuristics:*** *I feel the author is being a little alarmist but perhaps such urgency is needed to get people activated. I didn't see any real sources for the arguments he was making but it seems to line up with most of what I've read about the climate issue.*

> ***Content heuristics:*** *This article is provided by a trusted source organization. It also references specific verifiable facts and relevant issue organizations and individuals ….* ***Cognitive heuristic:*** *The content of this article also comports with my current understanding of the issue in question.*

The first comment above starts with a cognitive heuristic but then this prior belief shapes how the person views the rest of the article, judging it as alarmist, and paying attention to evidence that lines up with what the person had already known. The second argument uses the source heuristic and references used, which are also viewed in light of what the respondent already knows.

Taking a deeper look at the comments provided by respondents in the Emerging News condition we see that heuristics other than those associated with confirmation bias played a stronger role in shaping respondents' views of articles. The first two examples below are from respondents who disagreed with an article which we deemed as False and the third is from a respondent who agreed with an article which we deemed as True. All three respondents provided justifications that were rooted in the writing style, provision of supporting evidence, and credibility of the source, and to a much lesser extent, relying on prior knowledge and beliefs.

> *The article makes many claims but offers no evidence of these claims. The frequent unnecessary capital letters are a big hint, and so is the lack of sources or an author's first and last name. It is written poorly, and is meant to scare, not meant to inform. It just has an obvious fake tone to it, and makes ridiculous claims.*

> *The language used and style of writing give the impression the article was not written by a highly educated or scientific person. Although there are some technical words used, the overall impression of the article is "amateur".*

> *I believe the news article because it seems credible and factual. I know that there have been shortages of PPE, so it seems legitimate. Reuters is usually reliable and accurate, so I trust it.*

[4] Drilling deeper into these 72%, we note that some of the articles were more easily discernible as false and identified as such by a large percent of the respondents. When we examine specific articles that were less clearly distinguished as such, we see that proportions of agreement for these articles is at 43% for respondents in the No AI support. This proportion increases to 72% in the Generic AI condition, and 80% and 81% in the AI Source and AI Content conditions, respectively. Hence, the role played by the AI is significantly increased in these more confusing contexts, adding support to our hypothesis that AI will play a greater role when there are no strongly held prior beliefs.

Table 7
Heuristic Categories in Everyday vs. Emerging News.[51].

Do you believe the information in this news article?	Belief alignment	Experience alignment	Knowledge alignment	Supporting evidence	Bias	Accuracy	Coherent story	Writing style	Trusted source
Everyday News	29%	3%	22%	11%	12%	6%	6%	11%	14%
Emerging News	30%	1%	26%	13%	4%	11%	16%	11%	18%
Heuristic type	Self/Cognitive Heuristics			Content Heuristics					Source heuristic

Another indication of the lesser role of confirmation bias in this condition are comments from unsure respondents. For example:

I really do not know enough about chemistry or biology or whatever field of science this is to know if it is true or not. (respondent said they probably not believe the article that we marked as False)

There have been so many potential coronavirus rumers about ways to limit or cure the virus that I doubt them all. Until one has proven to be effective I will continue to be doubtful which is how I answered the question above. (respondent said they probably not believe the article that we marked as False)

I mean at this point anything is possible. (respondent said they might or might not believe the article that we marked as False)

Unsure respondents (those who chose "might or might not believe and article") accounted for 25% or the responses in the Emerging News and No AI condition, compared with only 16% in the Everyday News and No AI condition.

Still, in the Emerging News condition, cognitive heuristics were used, as our Table 7 shows, and we provide two examples below. But even here, justifications were not as elaborate as in the Everyday News condition, indicating that even when respondents relied on cognitive heuristics, their prior knowledge and beliefs were not very strong:

I believe in the information because it has been believed for many years that Vitamin C is amazing for colds and flus and it is what we take when we get sick. (respondent said they definitely believed the article that we marked as False)

I believe that the trials are happening and I have heard something about Vitamin C boosting the immune system. (respondent said they probably believed the article that we marked as False)

The first respondent above provides a general statement about Vitamin C in other contexts, the second respondent indicates they "have heard something" about the issue, and they state their belief in only one part of the article ("I believe that the trials are happening"). These justifications are generally weaker than those we have seen under the Everyday News condition.

6.3.2. Heuristics types and AI impact

As a final analysis, we looked at comments provided by participants that explicitly mentioned the AI. In the Everyday News condition, 8% explicitly mentioned AI and in the Emerging News condition, 15.4%. Within these justifications the AI was mentioned in addition to other heuristics. For example (our highlight):

The information is ridiculous and inaccurate. It is a silly conspiracy theory. I have never heard of the source and it has no credibility. It appears heavily biased and lacks any citation or scientific backing. ***The article is rated as a not trusted news source****. The whole premise is absurd.*

What was further interesting, was that while only a few participants explicitly mentioned the AI itself in their justification comments, the AI seemed to have triggered the use of relevant heuristics. For example, we found greater reliance on the source credibility as a heuristic in the AI Source treatment, and a greater reliance on the content heuristics, such as accuracy, in the AI Content treatment. This relationship was highly significant using a chi-square test under both the Everyday News and the Emerging News conditions. What this result implies is that when AI is tailored it has a significant impact on shaping the opinions of news consumers.

7. Discussion

We obtained multiple insights from study 2. First, hypothesis 1 about the greater effectiveness of AI advice in emerging news situation was supported, highlighting the need to act early in shaping correct judgements of news. Second, hypothesis 2 about the greater effectiveness of tailored (vs. generic) AI advice was also supported, albeit partially. Specifically, when respondents were open to accepting algorithmic advice, the tailored advice was more effective than the generic one. This again provides important insight to designers of news flagging tools with respect to how the advice should be provided.

Our qualitative analysis of the use of cognitive versus content and source heuristics, specifically in the context of confirmation bias, provides evidence to support the literature that when news consumers hold strong prior beliefs and are confident in those beliefs and in their ability to evaluate the news then the use of cognitive heuristics is significantly higher. This, in turn, reduces their willingness to "listen to" the advice of the algorithm. When respondents are unsure about a topic, they turn to other heuristics, namely content and source. In this case they are more willing to accept the AI advice.

The key insight from this qualitative exploration of comments is, again, that the timing of news veracity interventions is of high importance in determining their effectiveness, and that interventions should focus on emerging news situations. In those situations, providing specific explanation using either content or source heuristics is more effective than provide generic fake news flags and messages.

Our insights from study 2 emerge from its unique design, which provided us with rich data, both quantitative and qualitative. While our statistical tests are relatively simple, they are sufficient to test our hypotheses and elegantly demonstrate the impact of heuristics-based algorithmic advice. The rich comments that we obtain from the qualitative comments enable us to shed more light on this phenomenon of interest and to offer contribution to future theory development in this area.

8. Limitations

No study is without limitations. First, the random assignment of articles to respondents resulted in unbalanced data sets, which might have had impact on ease of assessment of news veracity. We don't believe this was a significant limitation given the random assignments and the type of analysis used in this paper (the chi-square and proportions tests) but further studies can build on more controlled lab experiment in testing our findings further. Second, our work only studies AI advice as a one-time phenomenon where readers do not have prior beliefs regarding reliability and expertise of this particular AI, based on previous repeated interactions. Furthermore, the AI in our study is presented outside an institutional umbrella that build it, which can impact attitudes of users towards the tool. Further research is needed to understand the impact of these additional factors. Finally, we did not study the potential impact of other AI interventions such as those based on writing style or

presentation of corrective or countering claims.

9. Contributions and future research

Our study makes several important contributions to both theory and practice. First, to the best of our knowledge, this study is the first study to explicitly consider the timing of news veracity interventions. As the literature we reviewed alludes to, the more exposure to a specific topic the higher potential for strongly shaped opinions and beliefs. In turn, it becomes more difficult to impact those beliefs, even in cases of clearly false news. Our work demonstrates that there is an important window of opportunity for providing algorithmic advice and for that advice being accepted. Future studies can build on this finding in multiple ways. For example, it is still not well understood how novel a news topic must be to have optimal interventions. In our study, we tested a wide range, between very novel COVID-19 pandemic and the reoccurring news topics of vaccination and climate change. However, more granular studies of this range can be done. There are also open questions about how the politicization of news topics, even if they are not inherently political topics, impacts the strength of beliefs and the effectiveness of interventions. Additionally, questions related to the long-term impacts of early veracity interventions still remain, such as how interventions during emerging news situations affects the development and spread of conspiracy theories and false information.

Second, we contribute to the literature on heuristics in decision making in two important ways. First, we identify the set of heuristics that are used in the specific news veracity context and shed light on how news consumers evaluate the news that they read. This can help media outlets present news to their readers in a more convincing ways, it can help in creating news tags and taxonomies, and it can help consumers to share news on social media more responsibly. Second, we demonstrate the tailored advice that builds on specific heuristics is more effective than generic advice. The role of heuristics in decision-making is well known, it seems that a misconception in algorithmic advice is that the advice itself can serve as a heuristic (i.e. "I believe it because the AI trusts it"). As it turns out, the AI in itself is not as strong of a heuristic as are attributes of the content and source of the article. Future research can narrow in on other specific heuristics and understand how to further tailor the advice of AI under different contingencies. Specifically, this study had participants read news articles with the intervention text atop the article. However, more often, news is consumed passively while scrolling through social media feeds. While it is possible that our interventions could be used as flags in a social media feed, this should be explicitly tested. Future research should not only continue to test heuristic tailored designs, but also test both active and passive consumption environments. In this context, interventions other than heuristics-based, which exist in the literature on AI (e.g. Swire & Ecker, 2018), should be further explored for the effect of the timing of interventions, which we found to be significant in this study.

Third, our qualitative analysis of heuristics used, with specific focus on the use of cognitive heuristics, offers contribution to the literature on confirmation bias formation and persistence in the news context. Our differentiation between the mere use of cognitive heuristics and the actual strength of those cognitive heuristics is an important one for better understanding how confirmation bias affects decision making. Specifically, we find that people generally rely on cognitive heuristics. In other words, they turn to seek validation of information within their prior knowledge and beliefs, regardless of context. However, in situations where such validation is weak, people begin to pay more attention to source and content heuristics, and those heuristics are more susceptible to AI advice. This insight fits well with the literature that cautions against trying to provide algorithmic advice that goes against one's strongly held beliefs.

Our paper also offers important contributions for practitioners, such as fact-checking organizations. Due to information overload, fact-checking organizations often fact check information that has already been highly engaged with. However, our findings suggest that efforts could be more effective if focus was shifted to novel, emerging topics. Hence, rather than filtering down information to check by how much engagement that information is currently receiving, it can be filtered down by how novel the topic of the information is, whether it is receiving engagement currently or not. Future work to build systems to assist fact checkers in finding these emerging topics, according to our findings, can be fruitful. Similarly, the findings of this work imply that the continued development of methods for early detection, warnings, nudges, or other information veracity interventions are of high importance.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research has been partially funded by National Science Foundation (United States) grant #1937143.

Appendix. Demographic and Control Variables

Table A1
Demographics of our Sample – Study 1

Demographics Age	Round 1 (n = 15)	Round 2 (n = 100)	Round 3 (n = 101)	Round 4 (n = 100)
18–24	0%	0%	2%	0%
25–34	20%	26%	24%	34%
35–44	33%	36%	29%	32%
45–54	27%	21%	27%	18%
55–64	20%	13%	13%	12%
65–74	0%	4%	6%	4%
Gender				
Female	40%	44%	51%	41%

(*continued on next column*)

[5] The numbers in the table add up to more than 100% because respondents can use multiple heuristics to justify their assessment.

Table A1 (*continued*)

Demographics Age			Round 1 (n = 15)	Round 2 (n = 100)	Round 3 (n = 101)	Round 4 (n = 100)
Male			60%	55%	48%	59%
Prefer not to answer			0%	1%	1%	0%
Highest level of education completed						
High school			13%	14%	9%	12%
Associate degree or some college education			54%	32%	28%	28%
Bachelor's degree in college (4-year)			33%	45%	50%	46%
Master's; Professional; Doctoral degree			0%	9%	14%	12%
Responses for each article type						
ID	Source	Stance	Count	Count	Count	Count
1	BBC	Pro-vax	1	6	13	10
2	Natural News	Anti-CC	0	9	5	6
3	NPR	Pro-CC	6	14	12	8
4	Freedom Bunker	*Anti*-vax	1	6	10	16
5	Jew World Order	Anti-CC	1	6	12	5
6	Chicago-Sun Times	Pro-vax	1	7	7	11
7	NPR	Pro-vax	2	9	7	11
8	The Gateway Pundit	Anti-CC	0	7	4	10
9	Collective Evolution	*Anti*-vax	2	8	6	5
10	Daily Kos	Pro-CC	0	12	3	8
11	Fortune	Pro-CC	1	8	12	4
12	Natural News	*Anti*-vax	0	8	10	6

Table A2
Controls in our Sample – Study 1

Social Media and News Consumption	Round 1 (n = 15)	Round 2 (n = 100)	Round 3 (n = 101)	Round 4 (n = 100)
Primary channel for news consumption				
Social Media	20%	27%	30%	26%
News Websites	73%	46%	42%	42%
TV	7%	21%	25%	26%
Newspaper	0%	3%	2%	4%
Other	0%	3%	2%	2%
57% of respondents indicated they use additional news sources beyond the abovementioned primary source. These sources included additional channels from the above list, as well as radio, friends and word of mouth, and other online sources (e.g. blogs, podcasts, content aggregators).				
How often do you consume news through this channel?				
Weekly	20%	16%	13%	3%
Multiple times a day	47%	35%	36%	46%
Daily	33%	48%	48%	47%
Less than once a week	0%	1%	4%	2%
Never	0%	0%	0%	2%
What social media do you typically use (select all that apply)?				
Facebook	23%	27%	27%	26%
Twitter	21%	16%	18%	17%
YouTube	19%	24%	22%	24%
Reddit	15%	16%	16%	20%
Snapchat	6%	3%	1%	2%
Instagram	8%	13%	12%	12%
Other	4%	1%	2%	1%
I do not use social media	4%	1%	1%	0%
When you use social media, how often do you share news?				
I do not use social media	13%	4%	6%	0%
Never share	20%	30%	29%	41%
Sometimes	60%	58%	60%	50%
Most of the time	7%	6%	4%	7%
Always share	0%	2%	1%	2%
What news sources do you usually trust (open ended question)? A broad range of responses include the following sources such as specific social media (e.g. Reddit), specific news "brands" (e.g. BBC, NBC, PBS, Fox News, NPR, etc.), specific newspapers titles, public radio, news organizations (e.g. Associated Press and Reuters), local news channels, alternative sources (e.g. Wikileaks and The Real News), Government websites, anecdotal evidence from people on the ground, right leaning sources, independent content creators on YouTube, and podcasts.				

Table A3
Demographics of our Sample – Study 2

Demographics Age	Everyday News (n = 651)	Emerging News (n = 477)
18–24	1%	3%
25–34	31%	38%
35–44	34%	32%

(*continued on next column*)

Table A3 (*continued*)

Demographics Age	Everyday News (n = 651)	Emerging News (n = 477)
45–54	20%	13%
55–64	11%	10%
65–74	4%	4%
Gender		
Female	43%	42%
Male	57%	58%
Prefer not to answer	0%	0%
Highest level of education completed		
High school	11%	9%
Associate degree or some college education	32%	31%
Bachelor's degree in college (4-year)	45%	45%
Master's; Professional; Doctoral degree	11%	13%

Table A4
Controls in our Sample – Study 2

Social Media and News Consumption	Everyday News (n = 651)	Emerging News (n = 477)
Primary channel for news consumption		
Social Media	31%	34%
News Websites	43%	42%
TV	21%	20%
Newspaper	3%	2%
Other	3%	2%
How often do you consume news through this channel?		
Weekly	12%	17%
Multiple times a day	36%	31%
Daily	49%	50%
Less than once a week	2%	2%
Never	1%	0%
What social media do you typically use (select all that apply)?		
Facebook	70%	73%
Twitter	54%	58%
YouTube	64%	71%
Reddit	46%	50%
Snapchat	6%	9%
Instagram	31%	32%
Other	3%	4%
I do not use social media	2%	1%
When you use social media, how often do you share news?		
I do not use social media	3%	1%
Never share	36%	35%
Sometimes	55%	59%
Most of the time	5%	4%
Always share	1%	1%
What news sources do you usually trust (open ended question)?		
A broad range of responses include the following sources such as specific social media (e.g. Reddit), specific news "brands" (e.g. CNN, NBC, PBS, Fox News, NPR, etc.), specific newspapers titles, public radio, news organizations (e.g. Associated Press and Reuters), local news channels, alternative sources (e.g. Wikileaks and The Real News), Government websites, anecdotal evidence from people on the ground, right leaning sources, independent content creators on YouTube, and podcasts.		

References

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *The Journal of Economic Perspectives, 31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Anspach, N. M. (2017). The new personal influence: How our facebook friends influence the news we read. *Political Communication, 34*(4), 590–606. https://doi.org/10.1080/10584609.2017.1316329

Aula, A. (2005). *Studying user strategies and characteristics for developing web search interfaces*. Tampere University Press.

Bago, B., Rand, D. G., & Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of Experimental Psychology: General, 1–18*. https://doi.org/10.1037/xge0000729

Baly, R., Karadzhov, G., Alexandrov, D., Glass, J., & Nakov, P. (2018). Predicting factuality of reporting and bias of news media sources. In *Proceedings of the 13th international workshop on semantic evaluation* (pp. 3528–3539). https://doi.org/10.18653/v1/D18-1389

Bates, J. A., & Lanza, B. A. (2013). Conducting psychology student research via the mechanical Turk crowdsourcing service. *North American Journal of Psychology, 15*(2).

Besedeš, T., Deck, C., Sarangi, S., & Shor, M. (2012). Age effects and heuristics in decision making. *The Review of Economics and Statistics, 94*(2), 580–595. https://doi.org/10.1162/rest_a_00174

Blumenthal-Barby, J. S. (2016). Biases and heuristics in decision making and their impact on autonomy. *The American Journal of Bioethics, 16*(5), 5–15. https://doi.org/10.1080/15265161.2016.1159750

Boczkowski, P. J. (2011). *News at work: Imitation in an age of information abundance*. The University of Chicago Press.

Boczkowski, P., Mitchelstein, E., & Matassi, M. (2017). Incidental news: How young people consume news on social media. In *Proceedings of the 50th Hawaii international conference on system sciences*. https://doi.org/10.24251/hicss.2017.217

Bodemer, N., Hanoch, Y., & Katsikopoulos, K. V. (2015). Heuristics: Foundations for a novel approach to medical decision making. *Internal and Emergency Medicine, 10*(2), 195–203. https://doi.org/10.1007/s11739-014-1143-y

Bozarth, L., & Budak, C. (2020). Toward a better performance evaluation framework for fake news classification. *Proceedings of the International AAAI Conference on Web and Social Media, 14*(1), 60–71.

Broockman, D., & Kalla, J. (2016). Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science, 352*(6282), 220–224. https://doi.org/10.1126/science.aad9713

Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. Guilford Press.

Chung, M. (2017). Not just numbers: The role of social media metrics in online news evaluations. *Computers in Human Behavior, 75*, 949–957. https://doi.org/10.1016/j.chb.2017.06.022

Colliander, J. (2019). "This is fake news": Investigating the role of conformity to other users' views when commenting on and spreading disinformation in social media. *Computers in Human Behavior, 97*, 202–215. https://doi.org/10.1016/j.chb.2019.03.032

Colman, A. M. (2015). *A dictionary of psychology*. Oxford University Press.

Cruz, A., Rocha, G., Sousa-Silva, R., & Cardoso, L. (2019). *Team fernando-pessa at SemEval-2019 task 4: Back to basics in hyperpartisan news detection*. https://doi.org/10.18653/v1/S19-2173

Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics?. In *Simple heuristics that make us smart* (pp. 97–118). Oxford University Press.

Davidson, D. (1995). The representativeness heuristic and the conjunction fallacy effect in children's decision making. *Merrill-Palmer Quarterly, 41*(3), 328–346.

Doswell, C. A. (2004). Weather forecasting by humans—heuristics and decision making. *Weather and Forecasting, 19*(6), 1115–1126. https://doi.org/10.1175/waf-821.1

Dvir-Gvirsman, S. (2019). I like what I see: Studying the influence of popularity cues on attention allocation and news selection. *Information, Communication & Society, 22*(2), 286–305. https://doi.org/10.1080/1369118x.2017.1379550

Ecker, U. K., Lewandowsky, S., Cheung, C. S., & Maybery, M. T. (2015). He did it! She did it! No, she did not! Multiple causal explanations and the continued influence of misinformation. *Journal of Memory and Language, 85*, 101–115.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.

Fischhoff, B., Kahneman, D., Slovic, P., & Tversky, A. (2002). For those condemned to study the past: Heuristics and biases in hindsight. *Foundations of cognitive psychology: Core readings*, 621–636.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*(4), 650–669. https://doi.org/10.1037//0033-295x.103.4.650

Gigerenzer, G., & Goldstein, D. G. (1999). Betting on one good reason: The take the best heuristic. In *Simple heuristics that make us smart* (pp. 75–95). Oxford University Press.

Go, E., Jung, E. H., & Wu, M. (2014). The effects of source cues on online news perception. *Computers in Human Behavior, 38*, 358–367. https://doi.org/10.1016/j.chb.2014.05.044

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review, 109*(1), 75–90. https://doi.org/10.1037/0033-295x.109.1.75

Grabe, M. E., Zhou, S., Lang, A., & Bolls, P. D. (2000). Packaging television news: The effects of tabloid on information processing and evaluative responses. *Journal of Broadcasting & Electronic Media, 44*(4), 581–598. https://doi.org/10.1207/s15506878jobem4404_4

Graefe, A., & Armstrong, J. S. (2012). Predicting elections from the most important issue: A test of the take-the-best heuristic. *Journal of Behavioral Decision Making, 25*(1), 41–48. https://doi.org/10.1002/bdm.710

Gruppi, M., Horne, B. D., & Adalı, S. (2020). *NELA-GT-2019: A large multi-labelled news dataset for the study of misinformation in news articles*. arXiv preprint arXiv:2003.08444.

Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., 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. *Proceedings of the National Academy of Sciences, 117*(27), 15536–15545. https://doi.org/10.1073/pnas.1920498117

Hassan, N., Nayak, A. K., Sable, V., Li, C., Tremayne, M., Zhang, G., Arslan, F., Caraballo, J., Jimenez, D., Gawsane, S., Hasan, S., Joseph, M., & Kulkarni, A. (2017). ClaimBuster. *Proceedings of the VLDB Endowment, 10*(12), 1945–1948. https://doi.org/10.14778/3137765.3137815

Hernandez, I., & Preston, J. L. (2013). Disfluency disrupts the confirmation bias. *Journal of Experimental Social Psychology, 49*(1), 178–182. https://doi.org/10.1016/j.jesp.2012.08.010

Hertzum, M., & Frøkjær, E. (1996). Browsing and querying in online documentation: A study of user interfaces and the interaction process. *ACM Transactions on Computer-Human Interaction, 3*(2), 136–161. https://doi.org/10.1145/230562.230570

Hoffrage, U., & Reimer, T. (2004). Models of bounded rationality: The approach of fast and frugal heuristics. *Management Revue, 15*(4), 437–459. https://doi.org/10.5771/0935-9915-2004-4-437

Holzinger, A., Biemann, C., Pattichis, C. S., & Kell, D. B. (2017). *What do we need to build explainable AI systems for the medical domain?*. arXiv preprint arXiv:1712.09923.

Horne, B., & Adali, S. (2017). This just in: Fake news packs A lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In *Eleventh international AAAI conference on web and social media* (pp. 759–766).

Horne, B. D., Dron, W., Khedr, S., & Adali, S. (2018a). Assessing the news landscape: A multi-module toolkit for evaluating the credibility of news. In *Companion of the the web conference 2018*. https://doi.org/10.1145/3184558.3186987. WWW '18.

Horne, B. D., Gruppi, M., & Adali, S. (2019, December). *Trustworthy misinformation mitigation with soft information nudging*. IEEE Xplore. https://doi.org/10.1109/TPS-ISA48467.2019.00039

Horne, B., Khedr, S., & Adali, S. (2018b). Sampling the news producers: A large news and feature data set for the study of the complex media landscape. *Proceedings of the Twelfth International AAAI Conference on Web and Social Media*, 518–527.

Horne, B. D., Nevo, D., O'Donovan, J., Cho, J. H., & Adalı, S. (2019, July). Rating reliability and bias in news articles: Does AI assistance help everyone? *Proceedings of the International AAAI Conference on Web and Social Media, 13*, 247–256, 01.

Horne, B. D., Nørregaard, J., & Adali, S. (2019c). Robust fake news detection over time and attack. *ACM Transactions on Intelligent Systems and Technology, 11*(1), 1–23. https://doi.org/10.1145/3363818

Horne, B. D., Nørregaard, J., & Adalı, S. (2019, July). Different spirals of sameness: A study of content sharing in mainstream and alternative media. *Proceedings of the International AAAI Conference on Web and Social Media, 13*, 257–266, 01.

Hosseinimotlagh, S., & Papalexakis, E. E. (2018). Unsupervised content-based identification of fake news articles with tensor decomposition ensembles. In *Proceedings of the workshop on misinformation and misbehavior mining on the web (MIS2)*.

Igartua, J. J., & Cheng, L. (2009). Moderating effect of group cue while processing news on immigration: Is the framing effect a heuristic process? *Journal of Communication, 59*(4), 726–749. https://doi.org/10.1111/j.1460-2466.2009.01454.x

Jang, S. M., & Kim, J. K. (2018). Third person effects of fake news: Fake news regulation and media literacy interventions. *Computers in Human Behavior, 80*, 295–302.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus And Giroux.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty : Heuristics and biases*. Cambridge University Press.

Kim, A., & Dennis, A. R. (2019). Says who? The effects of presentation format and source rating on fake news in social media. *MIS Quarterly, 43*(3), 1025–1039. https://doi.org/10.25300/misq/2019/15188

Kim, A., Moravec, P. L., & Dennis, A. R. (2019). Combating fake news on social media with source ratings: The effects of user and expert reputation ratings. *Journal of Management Information Systems, 36*(3), 931–968. https://doi.org/10.1080/07421222.2019.1628921

Knobloch-Westerwick, S., Sharma, N., Hansen, D. L., & Alter, S. (2005). Impact of popularity indications on readers' selective exposure to online news. *Journal of Broadcasting & Electronic Media, 49*(3), 296–313. https://doi.org/10.1207/s15506878jobem4903_3

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning & Memory, 6*(2), 107–118. https://doi.org/10.1037/0278-7393.6.2.107

Lau, R. R., & Redlawsk, D. P. (2001). Advantages and disadvantages of cognitive heuristics in political decision making. *American Journal of Political Science, 45*(4), 951–971. https://doi.org/10.2307/2669334

Lieder, F., Griffiths, T. L., Huys, M., J, Q., & Goodman, N. D. (2018). The anchoring bias reflects rational use of cognitive resources. *Psychonomic Bulletin & Review, 25*(1), 322–349. https://doi.org/10.3758/s13423-017-1286-8

MacGillivray, B. H. (2017). Characterising bias in regulatory risk and decision analysis: An analysis of heuristics applied in health technology appraisal, chemicals regulation, and climate change governance. *Environment International, 105*, 20–33. https://doi.org/10.1016/j.envint.2017.05.002

Marewski, J. N., & Gigerenzer, G. (2012). Heuristic decision making in medicine. *Dialogues in Clinical Neuroscience, 14*(1), 77–89.

Messing, S., & Westwood, S. J. (2014). Selective exposure in the age of social media. *Communication Research, 41*(8), 1042–1063. https://doi.org/10.1177/0093650212466406

Metzger, M. J., & Flanagin, A. J. (2013). Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of Pragmatics, 59*, 210–220. https://doi.org/10.1016/j.pragma.2013.07.012

Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. *Journal of Communication, 60*(3), 413–439. https://doi.org/10.1111/j.1460-2466.2010.01488.x

Minas, R. K., Potter, R. F., Dennis, A. R., Bartelt, V., & Bae, S. (2014). Putting on the thinking cap: Using NeuroIS to understand information processing biases in virtual teams. *Journal of Management Information Systems, 30*(4), 49–82. https://doi.org/10.2753/mis0742-1222300403

Moravec, P. L., Kim, A., & Dennis, A. R. (2020). *Appealing to sense and sensibility: System 1 and system 2 interventions for fake news on social media*. Information Systems Research. https://doi.org/10.1287/isre.2020.0927

Moravec, P., Minas, R., & Dennis, A. R. (2018). Fake news on social media: People believe what they want to believe when it makes No sense at all. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3269541

Myers, D. G., & DeWall, N. C. (2018). *Psychology* (12th ed.). New York: Worth Publishers.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175–220. https://doi.org/10.1037//1089-2680.2.2.175

Nørregaard, J., Horne, B. D., & Adalı, S. (2019). NELA-GT-2018: A large multi-labelled news dataset for the study of misinformation in news articles. *Proceedings of the International AAAI Conference on Web and Social Media, 13*, 630–638.

Park, J., Konana, P., Gu, B., Kumar, A., & Raghunathan, R. (2013). Information valuation and confirmation bias in virtual communities: Evidence from stock message boards. *Information Systems Research, 24*(4), 1050–1067. https://doi.org/10.1287/isre.2013.0492

Pennycook, G., Bear, A., Collins, E. T., & Rand, D. G. (2020). The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Management Science*. https://doi.org/10.1287/mnsc.2019.3478

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. Springer-Verlag.

Popat, K., Mukherjee, S., Yates, A., & Weikum, G. (2018). *Declare: Debunking fake news and false claims using evidence-aware deep learning*. arXiv preprint arXiv:1809.06416.

Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social Psychology, 34*(2), 243–281. https://doi.org/10.1111/j.1559-1816.2004.tb02547.x

Rachlinski, J. J. (2000). Heuristics and biases in the courts: Ignorance or adaptation. *Or. L. Rev., 79*, 61.

Renjilian, C. B., Womer, J. W., Carroll, K. W., Kang, T. I., & Feudtner, C. (2013). Parental explicit heuristics in decision-making for children with life-threatening illnesses. *Pediatrics, 131*(2), e566–e572. https://doi.org/10.1542/peds.2012-1957

Rollwage, M., Loosen, A., Hauser, T. U., Moran, R., Dolan, R. J., & Fleming, S. M. (2020). Confidence drives a neural confirmation bias. *Nature Communications, 11*(1). https://doi.org/10.1038/s41467-020-16278-6

Rubin, V. L., Conroy, N., Chen, Y., & Cornwell, S. (2016). June). Fake news or truth? Using satirical cues to detect potentially misleading news. In *Proceedings of the second workshop on computational approaches to deception detection* (pp. 7–17).

Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2020). FakeNewsNet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big Data, 8*(3), 171–188.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63*(2), 129–138. https://doi.org/10.1037/h0042769

Starbird, K., Arif, A., Wilson, T., Van Koevering, K., Yefimova, K., & Scarnecchia, D. (2018). Ecosystem or echo-system? Exploring content sharing across alternative media domains. *Proceedings of the Twelfth International AAAI Conference on Web and Social Media*, 365–374.

Starbird, K., Spiro, E. S., & Koltai, K. (2020, June 25). *Misinformation, crisis, and public health—Reviewing the literature, VV1.0. MediaWell.* Social Science Research Council. https://mediawell.ssrc.org/literature-reviews/misinformation-crisis-and-public-health/versions/v1-0/.

Sundar, S. S., Knobloch-Westerwick, S., & Hastall, M. R. (2007). News cues: Information scent and cognitive heuristics. *Journal of the American Society for Information Science and Technology, 58*(3), 366–378. https://doi.org/10.1002/asi.20511

Swire-Thompson, B., DeGutis, J., & Lazer, D. (2020). *Searching for the backfire effect: Measurement and design considerations.* https://doi.org/10.31234/osf.io/ba2kc In press but where?

Swire, B., & Ecker, U. K. (2018). Misinformation and its correction: Cognitive mechanisms and recommendations for mass communication. *Misinformation and Mass Audiences*, 195–211.

Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining "fake news. *Digital Journalism, 6*(2), 137–153. https://doi.org/10.1080/21670811.2017.1360143

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge : Improving decisions about health, wealth, and happiness*. Penguin Books.

Todd, P. M., & Gigerenzer, G. (2007). Environments that make us smart: Ecological rationality. *Current Directions in Psychological Science, 16*(3), 167–171. https://doi.org/10.1111/j.1467-8721.2007.00497.x

Wang, W. Y. (2017). *liar, liar pants on fire": A new benchmark dataset for fake news detection.* arXiv preprint arXiv:1705.00648.

Wang, L. X., Ramachandran, A., & Chaintreau, A. (2016, April). Measuring click and share dynamics on social media: A reproducible and validated approach. In *Tenth international AAAI conference on web and social media. Actually can't find this in the proceedings*.

Wang, D., Yang, Q., Abdul, A., & Lim, B. Y. (2019). Designing theory-driven user-centric explainable AI. *Proceedings of the 2019 CHI conference on human factors in computing systems - CHI '19*. https://doi.org/10.1145/3290605.3300831