Hindawi
Human Behavior and Emerging Technologies
Volume 2024, Article ID 1069670, 20 pages
https://doi.org/10.1155/2024/1069670

Review Article

Online Fake News Opinion Spread and Belief Change: A Systematic Review

Filipe Altoe,[1] **Catarina Moreira**,[2] **H. Sofia Pinto**,[1] **and Joaquim A. Jorge**[1]

[1]*INESC-ID/DEI-Instituto Superior Técnico-Universidade de Lisboa, Lisbon, Portugal*
[2]*Human Technology Institute, University of Technology, Sydney, Australia*

Correspondence should be addressed to Filipe Altoe; luis.altoe@tecnico.ulisboa.pt

Received 19 September 2023; Revised 22 February 2024; Accepted 21 March 2024; Published 30 April 2024

Academic Editor: Mirko Duradoni

Fake news has been linked to the rise of psychological disorders, the increased disbelief in science, and the erosion of democracy and freedom of speech. Online social networks are arguably the main vehicle of fake news spread. Educating online users with explanations is one way of preventing this spread. Understanding how online belief is formed and changed may offer a roadmap for such education. The literature includes surveys addressing online opinion formation and polarization; however, they usually address a single domain, such as politics, online marketing, health, and education, and do not make online belief change their primary focus. Unlike other studies, this work is the first to present a cross-domain systematic literature review of user studies, methodologies, and opinion model dimensions. It also includes the orthogonal polarization dimension, focusing on online belief change. We include peer-reviewed works published in 2020 and later found in four relevant scientific databases, excluding theoretical publications that did not offer validation through dataset experimentation or simulation. Bibliometric networks were constructed for better visualization, leading to the organization of the papers that passed the review criteria into a comprehensive taxonomy. Our findings show that a person's individuality is the most significant influential force in online belief change. We show that online arguments that balance facts with emotionally evoking content are more efficient in changing their beliefs. Polarization was shown to be cross-correlated among multiple subjects, with politics being the central polarization pole. Polarized online networks start as networks with high opinion segregation, evolve into subnetworks of consensus, and achieve polarization around social network influencers. Trust in the information source was demonstrated to be the chief psychological construct that drives online users to polarization. This shows that changing the beliefs of influencers may create a positive snowball effect in changing the beliefs of polarized online social network users. These findings lay the groundwork for further research on using personalized explanations to reduce the harmful effects of online fake news on social networks.

Keywords: fake news; influencers; online belief change; online opinion formation; polarization; social network sites

1. Introduction

Fake news has been linked to the rise of psychological disorders [1], the increased disbelief in science [2], and the erosion of democracy and freedom of speech [3]. Americans believe fake news to be a more severe problem than most of the other critical issues in the country [4]. Lately, deepfake technology [5] has been used in the production of videos of artificial intelligence (AI)–created avatars impersonating news anchors reporting on fake news [6]. The advancement of deepfake technology is adding to the realism of fake news artifacts, increasing the difficulty for news consumers to discern disinformation from real news. Generative AI is increasingly being used with malicious intent. There is evidence of scammers producing AI-generated fake voice messages to victims' family members, giving them dire news of being in immediate danger and requesting monetary help [7].

For these reasons, fake news is a hot research subject at the time of this writing [8]. The state-of-the-art is mainly targeting their detection. Initial approaches attempted to use machine learning classifiers over news content and

analyze the news source. However, these methods fell out of favor due to the need for a level of manual annotation that renders the approaches unusable in practical settings. Moreover, as the fake news spread in online platforms sometimes tends to be of viral velocity [9], these classifiers are not well suited to the task. The community identified this problem and has started automatically adopting deep learning techniques to extract features from online news posts. However, these black box models need more transparency to earn the news consumer's trust in their generated outputs. Detection explainability and visualization research started to gain momentum [10–12], leveraging explainable AI (XAI) [13]. Even though research on XAI has evolved significantly, it is still producing explanations that are too technical and suited to machine learning experts rather than the general audience [14].

More recently, the community seems to be developing an understanding that fake news online consumers' education may be as crucial as fake news detection [15]. In this context, education may require changing their opinions and their preconceived beliefs. Cold factual explanations defending an opposite position can sometimes backfire and further entrench news consumers' preconceived beliefs, especially in online social network site (SNS) groups where polarization is prevalent [16]. In response to this, the research community has started to explore approaches for changing consumers' nonfactual beliefs [12] that could be more efficient alternatives to the usual human-generated fact-checked explanation articles [17]. In [18], the authors argue that explanations that nudge readers into a reflective state are more efficient than purely factual ones in changing user beliefs of fake photographs. Analyzing emotions in speeches and consequent explanations using contrastive elements has also been investigated [19]. This motivated researchers to explore novel methods for creating more nuanced and emotionally resonant explanations to encourage people to reflect on their beliefs. Some studies have proposed generating artistic or emotional explanations as an alternative approach to changing user beliefs rather than relying solely on facts [20, 21]. These explanations are aimed at evoking an emotional response from the newsreader and gently nudging them into a reflective state.

Another work proposed a roadmap to personalizing fake news explanation systems [22]. Our general hypothesis is that fake news explanations personalized to some level and evoke an emotional response carry better odds of changing users' preconceived beliefs than purely factual explanations.

1.1. Computational Creativity (CC). CC is a subfield of AI research that focuses on computational systems that exhibit behaviors that unbiased observers could deem creative [23]. One of the most popular CC theories offered by the literature [24] has its foundations in what is known as the four Ps [25, 26].

The four P's theory categorizes the study of creativity from four vantage points: *Person*, or what about the agent makes them creative; *Process*, or what sort of actions are performed in the manufacturing of creative work; *Product*, or what about the output artifact is worthy of being called creative; and *Press*, or what about the cultural tendencies of the environment drive a given work to be deemed creative.

The product vantage point is of interest when the goal is to produce something useful to humans [27], applicable to the fake news explanation use case. The literature offers several examples of computer-generated creative artifacts, such as music parodies [28], memes [29], anecdotes, poetry [30], and jokes [31]. More recently, the introduction of large language models (LLMs) [32] opened up a new realm of possibilities around AI-generated creative artifacts. LLM fine-tuning is being researched as an approach to specialize LLMs in specific tasks that are better aligned with human-generated tasks [33]. Emotion-based personalized explanations can leverage this research to present explanations in a manner that best aligns with the user preferences and maximizes the value of the experience through emotion evocation. It has been shown that computers can generate artifacts that create an emotional impact [34]. Our longer-term research objective is to use CC-generated fake news explanations to verify our general hypothesis. Figure 1 illustrates the concept. Understanding the current approaches used in explanations to educate fake news readers and how online belief evolves, in general, will provide insight into the research opportunities for validating this idea.

The remainder of the paper is organized as follows. Section 2 highlights the intersections and differences in the current literature review works to our proposed scope. Section 3 presents and explains the research questions that motivate the review methodology. The methodology itself is included in Section 4. This section presents separate discussions on the rationale behind the inclusion and exclusion criteria, the search methodology applied, and a taxonomy organizing the included works in clusters to facilitate analysis. Section 5 presents the findings for each review domain. Section 6 discusses the reviewed papers, highlighting how the findings address the proposed research questions and present other identified patterns. This section also presents the identified grand challenges and related future work. Conclusions are summarized in Section 7.

2. Related Work

To the best of our knowledge, this is the first cross-domain systematic literature review that includes the models, methodologies, and user studies focusing on their influence on belief formation and how preconceived beliefs are changed in online and social network platforms. Furthermore, we pay special attention to the role that polarization, an orthogonal dimension to the three chief ones, plays in online belief change. This section reviews related work on each domain and highlights the current gaps our work fills.

2.1. Models. From a model's perspective, modeling of opinion dynamics has been an active object of study [35], driving the community to generate surveys of online opinion propagation [36] and trust propagation [37]. The author in [36] includes topics of interest to our work, such as stubborn agents, biased agents, and opinion manipulation, the work reviewed papers published before 2019. From a model's

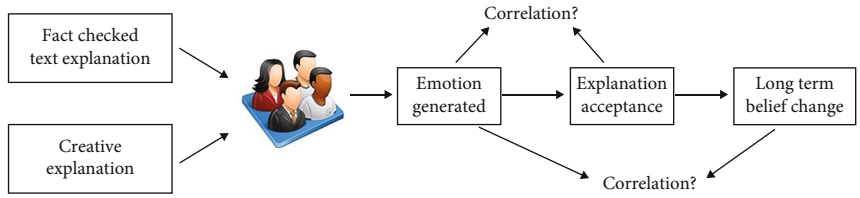

Figure 1: Belief change comparison: fact-based and CC-generated creative explanations.

perspective, we offer the community continuation of the work by Noorazar through the review of works published from 2020 and newer. Urena et al. [37] focus on how trust propagates in networks from a vantage point of opinions and recommendations.

Trust propagation in our context addresses how trust in specific opinion formation agents, such as influencers (INFLs), affects belief formation, a completely different approach.

Opinion formation models are another branch of opinion dynamics. Mastroeni et al. [38] specifically focus on agent-based models, which are centered on their mathematical formulation. Similarly, Abid et al. [39] also focus on the mathematical formulation of agent-based models. In contrast, we purposely exclude these works and only include the ones that have some practical validation through either simulation or dataset experimentation.

2.2. User Studies. From a user study perspective, the literature offers a few review works centered on specific online information domains, such as health [40, 41], politics [42], and online marketing [43, 44]. In the health domain, Wang et al. [40] executed a systematic literature review addressing misinformation spreading online health information. This study was performed before the COVID-19 pandemic, which makes it interesting from the standpoint of state-of-the-art before the event that has dominated health misinformation studies since 2020. While the methodology applied in the work was thorough and the findings around misinformation in online health insightful, the work does not address the vantage point of online belief change. The literature also offers works on COVID-19-related disinformation. Conspiracy theories are directly related to belief formation and opinion spread. Different conspiracy theories were born during the first year of the COVID-19 pandemic. Tsamakis et al. [41] performed a systematic literature review on COVID-19-related conspiracy theories. It focused on their prevalence, determinants, and public health consequences. An interesting result presented by the work, albeit somewhat predictable, was the higher prevalence of politically motivated COVID-19 conspiracy theories than other determinants. The work presented studies in the dimensions of demographics, level of income, psychological factors, religion, political orientation, and trust in science. However, it marginally addressed online beliefs related to acceptance of this class of conspiracy theories, a topic covered by our work.

Politics is another online belief-related research hotbed. One particular vantage point is online political participation (OPP). The level of OPP has been linked to disinformation and conspiracy theories [45]. Therefore, the topic is relevant from an online belief standpoint. The literature offers a systematic literature review of definitions and measurements of OPP [46]. The finding most relevant to online belief is that OPP is not an online equivalent of traditional offline political participation. It is instead shaped by and contingent upon the online platform on which the participation is conducted. The work does not elaborate further on the specific characteristics of platforms that enforce political beliefs that influence OPP. We hope our survey will provide further insight into this topic.

Opinion formation is also important in the domain of online marketing. Specific to online marketing, product, and service reviews and ratings are driving forces of online opinion formation. A systematic literature review and comparative study on how reviews and ratings influence opinions on buying and usage of the products were presented in [43]. The work concludes that regular consumer reviews are more influential in opinion formation than recommendations by professionals and paid experts. It may be seen as a use-case example of a social influence–based opinion model for the online marketing domain. Our work will attempt to find approaches that can be applied across domains.

2.3. Methodologies. From the methodology dimension perspective, the authors of [47] review belief dynamics processes from psychology, sociology, economics, philosophy, biology, computer science, and statistical physics perspectives. The work proposes a framework to enable comparisons of different belief-capturing methodologies. Even though individual belief is included as a structural component of the framework and is briefly discussed, the framework limits its modeling to a typical statistical physics approach. Our work is aimed at a more holistic view of the existing methodologies and studies the ones best suited to capture belief change. In our work, we felt that it is appropriate to combine models and methodologies into a single section named *opinion dynamics*, presented in Section 5.3.

2.4. Polarization. Lastly, for the polarization dimension, a notable systematic review links social media to polarization, synthesizing the contingent factors and underlying processes [48]. The work provides three aspects of polarization conceptualization: the ideological or opinion-based concept, the affective concept of disliking people from outgroups, and the social concept of avoiding the company or linkage with outgroups. This leads to presenting a conceptual framework of social media and polarization. Another study reviews

political polarization from a psychology vantage point [49]. The work provides a functional conceptualization of polarization in an attempt to explain how polarization may occur across partisan fault lines. It provides arguments that polarization is most likely to occur in scenarios of belief conflicts in society, such as in politics. Situations of belief conflicts tend to drive the formation of opposed belief groups, which are prone to polarization. Even though these works provide a rich link between online beliefs and polarization, neither studies the effect that polarization may have on constraining belief change. Our findings from that vantage point are presented in Section 5.2.

3. Research Questions

Figure 1 illustrates the conceptual idea for the future research setup to validate our main hypothesis. However, several questions remain unanswered regarding the detailed implementation of this concept. The research questions presented in this section were designed to help answer some of these questions.

3.1. RQ1—What Are the Main Drivers of Online Belief Change? Understanding the primary motivators that lead online users to change their preconceived beliefs is pivotal for designing personalized explanations that efficiently educate users in the event of fake news beliefs. This understanding is also central to the design of experiments to validate the main hypothesis of this work.

3.2. RQ2—How Are Current Opinion Models Being Used to Capture Belief Changes? Opinion dynamics covers a wide range of social science phenomena, such as the appearance of fads, consensus building, collective decision-making, rumor spreading, extremist expansion, and even cult propagation [50]. This RQ constrains the analysis of the models to focus on online belief change. RQ2 intends to understand if there are specific models offered by the literature that can be applied to collected experimental data to facilitate the identification of belief change by the participants.

3.3. RQ3—What Role Does Polarization Play in Changing Online Users' Preconceived Beliefs? This RQ explores the direct effect of information bias and polarization in changing online users' beliefs. This is important since the chief objective of fake news systems is to align news consumers with factual news. For that to happen, people with preconceived beliefs in fake news shall be shown that their beliefs are not based on facts and ought to change. Understanding not only if polarization is an important force potentially preventing belief change but also if there are documented approaches to best deal with this driver may provide insight into the explanation content and presentation that carry the best odds of success. Furthermore, it may add a dimension to the experiment as we can compare the polarization effect on purely factual and personalized explanations.

3.4. RQ4—What Alternative Approaches to Offering Fact-Checked Explanations Have Been Pursued by the Literature in an Attempt to Change Preconceived Online Beliefs? This RQ is aimed at identifying whether the literature offers alternative explanation methods beyond the usual fact-checked textual ones. The discovery of alternate explanations may modify and expand the high-level experiment design in Figure 1 into other dimensions of comparison between creative explanations and other existing types.

3.5. RQ5—How Can the Identified Belief-Changing Approaches Be Generalized to Multiple Belief Domains? This RQ will seek insights into whether any of the existing belief-changing approaches have the potential to be generalized into a framework that can cover multiple belief domains. We will highlight their strengths and weaknesses from a generalization potential standpoint. We will conclude the analysis by providing recommendations toward the generalization goal.

4. Methodology

This section presents the methodology used to create the final corpus of work reviewed. It explains the inclusion/exclusion criteria, the methodology applied, and the taxonomy used to classify the reviewed works.

4.1. Inclusion and Exclusion Criteria. This work includes peer-reviewed journal articles and conference papers of more than four pages in length, published in 2020 and later found in the following databases: Scopus, ACM, IEEE Xplore, and Web of Science. Nonpeer-reviewed articles and book chapter papers not centered on online belief and opinion change are excluded from the review. Opinion dynamics theory works are included when validated via dataset or simulation-based experiments. Purely theoretical papers are excluded. These papers are deemed too far removed from the goal of hypothesis verification as they would still need to be validated through experimentation. One of this work's goals is to understand potential psychological and social forces that may constrain acceptance of fact-based explanations. Therefore, polarization papers focusing solely on algorithm bias effects on polarization are excluded. Table 1 summarizes the inclusion and exclusion criteria.

An initial search according to the following keywords (online OR "social network") AND ("opinion change" OR "belief change" OR "change in belief" OR "opinion formation") filtering peer-reviewed papers published in 2020 or sooner returned 372 hits. A visualization created with the VOSviewer software application [51] was performed to identify potential clusters. Figure 2 presents the initial visualization results. Four high-level clusters were identified: user studies (green), opinion dynamics related to opinion formation (red), belief related to user intervention (purple), and public opinion (blue). Furthermore, the user studies cluster revealed the specific domains of COVID-19, climate change, education, and politics.

There are some noticeable correlations between the visualization of Figure 2 and the taxonomy presented in Section 4.2. We originally named the red cluster "opinion formation" due to the highest prevalence of the word "formation" in the visualization. However, the reading of the works revealed that the highest prevalence of the word was not

Table 1: Inclusion and exclusion criteria.

Inclusion	Exclusion
Peer-reviewed journal articles and conference papers longer than four pages in length	Workshops, book chapters, surveys, and nonpeer-reviewed papers
2020 and newer	2019 and older
Opinion dynamics modeling belief change validated by experimental data	Theoretical models only. User studies and simulated models not affecting belief change or formation
Polarization affecting beliefs	Algorithm bias effect on polarization

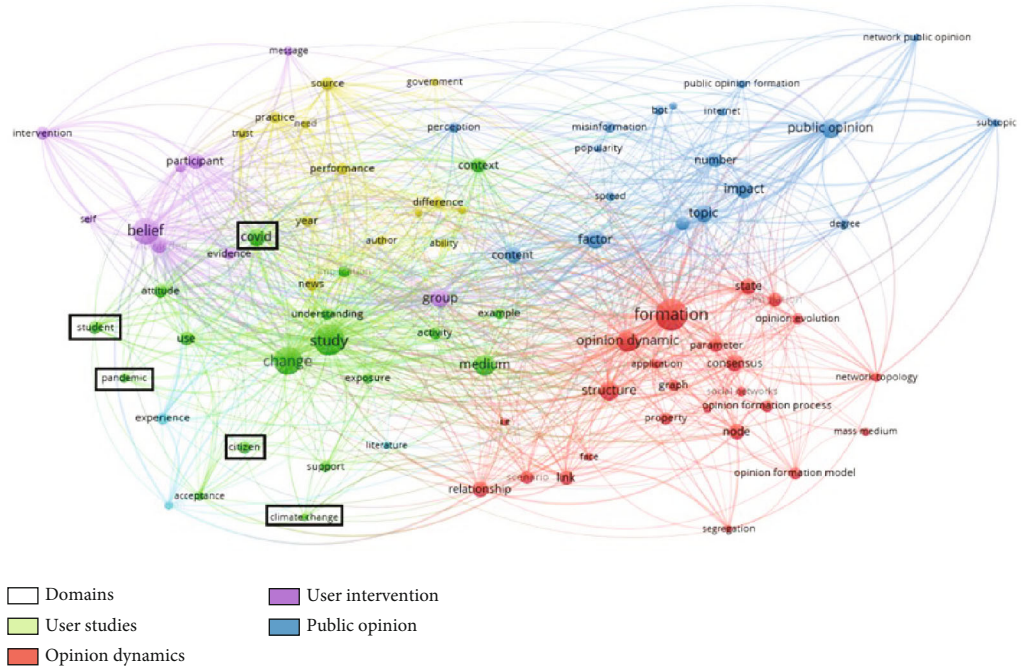

Figure 2: Opinion and belief change 2020–2023.

necessarily due to the cluster being about opinion formation but because of its relationship with the "opinion dynamics" subject. This is why the word "relationship" is also highly prevalent in the obtained visualization.

Terms with lower prevalence than opinion dynamics but with significance, such as "opinion formation process" and "opinion formation model," are, in fact, opinion dynamics implementations either through simulation or dataset experimentation. This is the reason that the opinion dynamics dimension is subdivided into simulation and dataset experiment clusters. Other terms such as "structure," "opinion evolution," and "network topology" represent applications of opinion dynamics techniques. The techniques can either focus on network structure (NETS) or the dynamics of opinion evolution in different contexts. These contexts are, namely, situations of crisis (CRIS) or traumatic events, the influence of stubborn or strong opinioned neighbors and INFLs, analysis of confirmation bias or homophily (HOMY), analysis of sociological or psychological forces in opinion dynamics, and the study of group or public opinion formation. All of these contexts became orthogonal clusters of the taxonomy as they are relevant to all dimensions of the review.

The blue cluster shows the term "public opinion" as highly prevalent because of its orthogonality with all three taxonomy dimensions. In this context, "public opinion," "public opinion formation," and "network public opinion" were combined in a single cluster: group opinion (GRPO). The high prevalence of "impact," "topic," and "factor" is somewhat synonymous in our context. The green cluster groups the user study papers. The identified user study domains in the visualization, COVID-19, student, citizen, and climate change became taxonomy clusters. The terms "support" and "exposure" were related to the sex and homosexuality theme, which was turned into a taxonomy cluster.

The visualization motivated adding other keyword searches centered on the following topics: belief and opinion changing user studies, belief formation models and methodologies, opinion formation models and methodologies, and polarization. The set of keywords for each search is illustrated in Figure 3.

Applying these criteria to the corpus yielded a total of 91 papers that were reviewed. These papers received a combined 782 citations at the time of this writing. Sixty-six of these works were published in journals, and the remainder

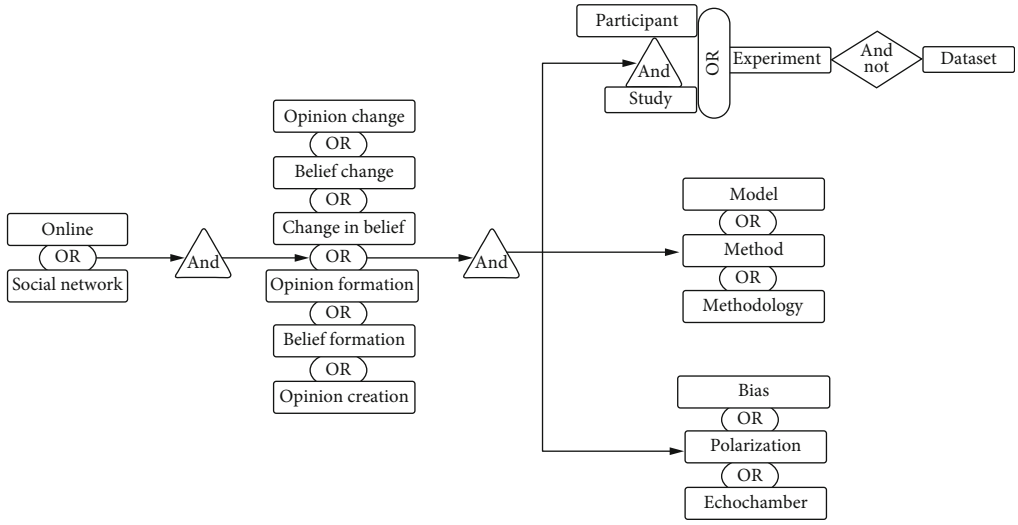

Figure 3: Keyword combinations.

in conferences. Fifty-one percent was published in Q1 journals and 10% in Q2 journals. Seven percent was published in A conferences and 9% in B conferences.

4.2. Taxonomy. The application of the methodology presented in Section 4 drove the organization of this review into three dimensions: domain specific, opinion models, and polarization.

Table 2 shows the taxonomy classification and corresponding works assigned to clusters. Some papers appear in more than one cluster. The description of each cluster is presented below.

4.2.1. TRMA (Trauma). It includes papers addressing belief change as a result of traumatic events. TRMA has been defined as "the experience of a vital discrepancy between threatening factors in a situation and individual coping abilities" [138]. TRMA can be objective and subjective [139]. Objective traumatic TRMA directly leads to post-traumatic stress disorder (PTSD). Subjective traumatic events may or may not.

4.2.2. INFL. It includes the effect of INFL [140] agents in belief change.

4.2.3. HOMY. HOMY is attributed to people's natural tendency to associate with people similar to themselves. Studies have documented that even infants as young as 6 months of age already show HOMY [141].

4.2.4. BCHB (Biaschamber). We dubbed BCHB as a combination of echo chamber and confirmation bias. An echo chamber is defined as the formation of like-minded online users reinforcing a narrative [142]. Confirmation bias is defined as the seeking to interpret evidence in ways that are partial to existing beliefs [143].

4.2.5. PSOC (Psychosocial). Even though HOMY and BCHB are PSOC phenomena, this cluster includes other sociological and psychological constructs.

4.2.6. GRPO. It focus on belief change of groups and public opinion.

4.2.7. SNSBs (SNS Biases). It focuses on SNSBs, which include filter bubbles [144] and other bias-inducing algorithms used by social networking sites.

4.2.8. NETS. It focuses on the influence that neighbor agents may have on belief change for groups within the same network.

4.2.9. CRIS. It includes CRISs that did not lead to TRMA.

4.2.10. STROs (Strong Opinions). It includes papers that address agents with STROs about a subject, stubborn, and zealot agents.

5. Findings

This section presents the survey findings in the context of its taxonomy.

5.1. Domain-Specific Dimension. As shown in Table 2, the *domain-specific dimension of the review was split into five classes*: *COVID-19*, *climate change*, *education*, *politics and policy*, and *other*. This section presents the findings for this dimension. Trust can be defined as "a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another" [145]. Trust is, therefore, a psychological construct, and it was one of the central drivers of opinion formation [79, 80] and belief change for the works reviewed in this dimension [52]. Trust in celebrities, namely, parasocial relationships [65], and social network INFLs was exploration topics. It was seen that INFLs can significantly affect people's opinions on different issues [52, 54] and that trust in the information source is an essential driver of belief change. Trust in government and officials was correlated with the consistency of the public messaging, affecting online belief change [70].

Table 2: Taxonomy of included papers in review.

	TRMA	INFL	HOMY	BCHB	PSOC	GRPO	SNSB	NETS	CRIS	STRO
Domain specific										
COVID-19		[52–54]		[55]	[56]	[54–56]				[55]
Climate change					[57]			[58]	[59]	
Education		[60]			[61, 62]	[60]			[60]	[63]
Politics and policy	[64]	[65, 66]	[67]		[65, 68, 69]	[70]		[66]	[69]	[64]
Sex and homosexuality					[71, 72]					
Other	[73]	[18, 74]		[75]	[74, 76–78]	[79]	[75]	[79, 80]		
Opinion dynamics										
Simulation		[81–83]	[84]		[85]	[83, 86]		[83, 85–93]	[94]	[95–97]
Dataset experiments		[98–100]	[101, 102]		[102–106]	[86, 98, 100, 103, 107, 108]		[86, 100]	[107]	
Polarization										
Theoretical studies			[102]	[102]	[102]		[109]			[110]
Models		[111, 112]	[113]	[114, 115]	[16, 116]	[117, 118]	[119–122]	[111, 123, 124]	[125, 126]	
User studies	[127]		[128]	[129, 130]	[131–133]	[134, 135]	[136]		[127, 137]	

Other psychological constructs were addressed in this dimension. Normative influence [68] is popularly known as herd mentality. Normative influence originates with the basic human desire to not stand apart from a group, i.e., the desire for social acceptance, which varies according to one's perceived risk of social rejection [146]. Informational influence is defined as the use of group knowledge as a determinant of correct beliefs [147]. In [68], the authors showed that factual information was not the primary driver of voting behavior. Normative influence may be powerful enough to influence belief change toward conformity to group decisions, even in secret voting. HOMY is another. In [67], the authors showed that it is possible to build a profile of former US President Donald Trump's supporters using HOMY as one of the most predictive signals for the model. Political scientist Elizabeth Noelle–Neumann proposed a theory dubbed the spiral of silence [148], highlighting individuals' unwillingness to publicly express an opinion when they feel that it may be against the majority opinion. This was verified through the opinion of social media users on LGBT acceptance in Nigeria [72, 79]. Bandwagon effects refer to individuals' tendencies to conform to predecessors' decisions [149]. In an online scenario, users are likely to rely on and gravitate toward more popular opinions as a form of mental shortcut. The authors of [74] showed strong evidence of further polarization of preconceived beliefs away from the expert's presented opinion. TRMA is another psychological event that can drive core belief change [150], with results showing that people who underwent intense TRMA feel that they changed their beliefs toward humanity. Post-traumatic growth (PTG) is defined as how individuals can experience positive psychological change after a traumatic event [151]. The authors of [73] showed deliberate rumination [152] to mediate the relationship between core belief challenge and PTG. Still, in psychology, the authors of [76] attempted to understand whether specific personalities are more prone to be persuaded into changing their beliefs, but their results were inconclusive.

It was observed that the polarization of subjects seems to be cross-correlated. Political affiliation was shown to be a polarization topic to be at the center of cross-correlation with topics such as environmentalism and climate ideology, COVID-19 response, policy preferences, immigration and patriotism, and even beliefs in biological attribution to homosexuality [71], which is directly related to support for homosexual rights. The results obtained in [54] revealed a connection between political viewpoints and misinformation regarding hydroxychloroquine (HCQ) in treating COVID-19 despite not being supported by scientific evidence. The author in [69] showed how political orientation is critical in shaping how public crises are interpreted and how belief changes about them. The authors of [58] show an association between left/right political ideology and environmentalist/skeptic climate ideology, respectively.

The connectivity between the traumatic public events and the arousal of emotional processes was demonstrated. Examples are the historical and institutional racism added to historical TRMAs such as the Tuskegee syphilis study [153] and the unethical and nonconsensual use of cancer cells from Henrietta Lacks [154] providing context for understanding vaccine hesitancy among Black individuals and their distrust of healthcare professionals and researchers [52]. A CRIS also evokes emotional responses that lead to polarization. This was demonstrated in [59] for the climate change topic and the topic of public opinion about police funding [64] at around the time of the murder of George

Floyd at the hands of law enforcement in the United States [155]. It was also seen that the level of a person's stubbornness was shown to be inversely correlated to the probability of belief change [55]. The hypothesized correlation between emotional arousal and polarization was confirmed.

Furthermore, a strong and direct stance stating the content is fake invariably leads to conflict, aligning with the finding that presenting factual explanations defending an opposite position can sometimes backfire and further entrench polarized people in their preconceived beliefs [16]. Stubbornness can lead to the entrenchment of beliefs. This can be seen even in less polarized topics, such as primary school teachers' beliefs about teaching computer science [63]. The results of this work showed that younger, less experienced teachers showed no signs of belief perseverance. Conversely, older, more experienced teachers demonstrated higher levels of belief perseverance, even when they indicated positive reactions toward the received computer science training.

Multiple studies provided evidence of the efficacy of explanations that nudge people into a state of reflection about their preconceived beliefs [18]. Ruffin et al. argue that attempting to explain how fake photographs were manipulated offers better results if done cautiously [18]. The authors of [62] showed this to be also valid in the context of belief change related to the nature of intelligence. Their results showed that rather than convincing people that intelligence is malleable, gentle mindset interventions may be the most important activity for helping them reflect on intelligence's malleability. This nudging may happen with the help of an emotion-evoking explanation, for example. Emotion responses were correlated to low knowledge in the process of a layperson acceptance and resultant opinion formation related to climate engineering approaches [57] and the driving of belief change of teachers under online and blended delivery methods [61]. The results showed that increasing knowledge about the topic in both cases drove belief change. This validates the concept that if knowledge is low regarding a given topic, emotional responses are used as indicators for attitudes toward or against a stimulus [156]. It also reinforces the need to balance an emotion-evoking explanation with facts to drive an increase in subject knowledge.

Explanation personalization was also addressed in this dimension. The authors of [77] showed that a personalized online algorithm–based intervention can change beliefs that may lead to inappropriate antibiotic demand by patients. Conversely, results obtained in [75] show that personalization enhances user experience, but the so-called "filter bubbles" favor the emergence of opinion polarization and radicalization through confirmation bias. One final noteworthy comment is about an interesting approach using sentiment analysis (SENTANL) pre- and postevent to capture belief change [56]. This methodology is promising and should be investigated further as a potential approach to validate this work's central hypothesis.

5.2. Polarization Dimension. As shown in Table 2, the review's polarization dimension was split into three classes: theoretical studies, models, and user studies. This section presents the findings for this dimension.

PSOC polarization driving forces were identified, namely, normative influence [116, 157], spiral of silence [148], confirmation bias, backfire effect, parasocial relationships, and HOMY. Confirmation bias influences polarization as the intensity of preconceived beliefs is sometimes the controlling aspect of belief change [130]. Arguably, people seek communities with higher chances to confirm their beliefs [136]. The results obtained in [16] showed confirmation bias in combination with the backfire effect to be strong drivers of polarization. The authors in [133] showed evidence of polarization development in another combination of PSOC constructs: parasocial relationships and HOMY. It was shown that people became further entrenched in their preconceived beliefs in the case of a contradicting opinion from a subject matter expert celebrity. In [131], the authors showed that feelings of resentment were the most significant predictor of the Black Lives Matter movement's support. Low-resentment individuals who expressed themselves on social media more frequently were less supportive.

Some papers demonstrated how some fragmented networks self-organize into multiple echo chambers of consensus and that consensus is a precondition for the emergence of polarization [109, 113, 114, 126, 128]. The authors of [115] looked even further into the correlation between echo chambers and polarization. The authors argued that their results validated the idea that echo chambers create a stable environment of confirmation bias and can even actively alienate some group members from outside contradicting information sources [158]. Similar results were obtained in [117, 129]. Another relevant finding was that if the same argument is presented by two people, one from their community and the other from another network, the likelihood of acceptance of the former is notably higher. This suggests that one possible way to reduce polarization may be to change beliefs from within. Focusing on changing the beliefs of key members, such as INFLs, of a polarized group may trigger a snowball effect in the beliefs of all members of the given community. The results in [123] suggest that this may be the case as they showed that most individuals from a network over time switch to opposite sentiments about the preconceived belief. The results obtained in [111] suggest that another possible way to revert polarization is to shield the members from their corresponding echo chambers, allowing them to access the ideas of members outside these chambers freely.

Evidence also acknowledges that user adherence to misinformation may sometimes be shifted away from accuracy and toward other goals. In [102], the authors concluded that providing subtle accuracy nudges is a promising approach to improving the quality of shared news. The correlation between SNSB and polarization was analyzed and verified [119]. Arguably, there is also a correlation between SNSB and individual PSOC constructs. Correlations between polarization due to the spiral of silence [121] and filter bubbles [122, 159] were demonstrated when people are influenced by strong SNSB. Another study looked at SNSB

through the lens of how people change their opinions when exposed to viral content [120]. The results showed that polarization barely increased after a regular marketing campaign and significantly increased upon the spread of polarized content.

The cross-correlation between polarized subjects also becomes evident after the review of this domain. It seems that political ideology is the central topic of polarization, and it can become cross-correlated with other polarization-prone subjects such as minority equality [137], patriotism, welfare policy [135], and the response to health crises [134]. The authors of [132] demonstrated that this cross-correlation directly correlates with emotion. They concluded that a psychological factor that impedes climate change beliefs is not related to climate but is mainly motivated by the feelings of dislike one political group feels toward the opposing group. CRIS events were also connected to the emergence of polarization [127, 137].

5.3. Opinion Dynamics Dimension. This section presents the findings for the *opinion dynamics* dimension. It includes the research works split into two classes, simulation and dataset experiment, as defined in Table 2.

The study of the effects of PSOC constructs was also present. HOMY is an important one. The results in [104] showed the effects of HOMY in the formation of echo chambers. It also demonstrated a moderate to high resemblance of the echo-chamber phenomenon for network topologies of abortion, capitalism, and feminism. This aligns with trends from other dimensions, suggesting cross-correlation between polarization topics. In [106], the authors show a context of evolving HOMY in political social network interactions. The results in [83] showed how HOMY and the spiral of silence drive people to form online social groups. The bandwagon effect, or herd mentality, influences consensus formation, as verified in [149]. The author in [101] showed a tendency for moderate online users to move toward the average opinion of their online friends. The authors in [98] showed that the bandwagon effect has a stronger driving force than INFLs and that the reach of consensus will be magnified in a scenario of bandwagon effect. However, this does not happen in highly segregated opinion networks [93]. This is an important finding as it suggests that polarization can be avoided if education on fake news posts happens at the initial stages of a social network before its consequent evolution to consensus. It was demonstrated in [81] that it is more difficult for someone to reach a consensus with a person who belongs to a group with a higher proportion of low-educated people than with a higher proportion of high-educated people. Another data point that shows the importance of educating online users on fake news posts. PSOC constructs are part of what forms a person's individuality. Individuality is also important regarding how personal experiences help shape GRPOs. The authors in [103] showed how GRPO results from the community's combined individual experiences. The authors argue that the so-called expert agents, or agents that bring strong individual experiences aligned with subjects of interest to the group, are highly influential to group beliefs.

Some studies highlighted the importance of a solid factual foundation to balance emotional arousal that nudges people onto reflective states for a higher probability of changing polarized beliefs. Emotion was confirmed to be an important component in this nudging, especially when balanced with other cognitive functions. Emotion was shown to be correlated with the higher interest people showed in resharing audio messages than purely text messages on social networks [102]. We hypothesize that audio messages have the potential to carry more emotional content than textual messages, driving people to have more interest in resharing them. The results in [85] showed that an online post combining affective and cognitive content increases people's willingness to share the message. Conversely, effectively weak and mostly cognitive content was shared the least. The nudging also needs to be founded on facts. The authors in [86] showed how removing facts from a post alienates people, and this alienation drives the emergency of nonfactual subtopics. The formed new subgroups can lead to a phenomenon known as information gerrymandering [160], where STRO individuals can keep negatively held opinions alive, even if nonfactual, as demonstrated in [96].

Information alienation was highly correlated with the emergence of polarized subnetworks. Therefore, it is important to share information about a given topic of interest to public opinion as early as possible, especially during a CRIS [94]. However, this needs to be done carefully to avoid a scenario of inconsistent messaging in case the results need to be reviewed later. The information revision may cause a backfire effect as [70] has provided evidence that inconsistent messaging reduces the effectiveness of explanations targeted at changing group beliefs. The constant changing of messaging was shown to generate a breach of trust by the public concerning the source of the message.

Research from this dimension found evidence that large networks with a diversion of opinions evolve into several smaller networks where consensus is reached and then polarization develops [87, 89]. However, Mansouri and Taghiyareh [82] show that when influential leaders exist in a social network, segregation has less impact on opinion formation than the effect created by INFLs. This shows how INFLs are key drivers of belief change in opinion networks [97], including public opinion formation [92]. This effect was also verified when mass media played the role of INFLs [84]. It was shown that even a small percentage of INFL-type agents motivated to manipulate opinion toward a specific goal could shape the majority opinion [100]. Similar results were shown in [99].

Being the intermediate step between opinion segregation and polarization, consensus needs to be understood. The results in [95] reveal that consensus in a multitopic network can be achieved if the number of stubborn agents around the subjects is small. Lastly, natural language processing (NLP) SENTANL on social network posts to identify belief change was also present in this domain [108]. This seems to be the preferred technique for identifying belief change by online users and is used across application domains.

6. Discussion

Section 5 presented the findings for the three dimensions included in this review. It presented the trends found in each dimension separately. As the chief focus of this work is a holistic review of online belief change, we consider the trends that appeared in more than one of the investigated dimensions to be the most relevant to belief change. This section summarizes these trends, and Table 3 shows the number of reviewed papers that addressed each of them in their corresponding dimensions. A given reviewed paper may have included more than a single trend. Conversely, some reviewed papers may have addressed a trend that did not appear in multiple dimensions. The following acronyms define each cross-dimensional trend.

- *SUBCNS (subconsensus):* Several works reviewed showed that an initially segregated network naturally self-organizes into multiple echo chamber subnetworks of consensus and then further evolves into corresponding polarization groups.
- *FACBAL (factual balance):* Multiple works showed that a combination of knowledge increases positive emotion, gently nudging online users into a reflective state. It became clear the importance of explanations to balance facts and emotional evocation for increased odds of success in belief change.
- *INFL:* INFL in SNS was confirmed to be a high driving force of belief change.
- *SENTANL:* NLP SENTANL before and after a specific event seems to be the preferred technique for identifying belief changes by online users.
- *PSOC:* This cross-dimensional trend includes the PSOC constructs linked to belief change. Psychosociological tendencies such as filter bubbles, spiral of silence, confirmation bias, backfire effect, parasocial relationships, and HOMY facilitate the evolution of social network groups into a scenario of polarization. Trust was also shown to be a psychological construct, arguably one of the strongest drivers of belief change in this group.
- *CRSPOL (cross-polarization):* All three dimensions showed that subjects' polarization is cross-correlated in different subjects, with political ideals being the central polarization topic.
- *ERLPUB (early publication):* Several works showed the importance of publishing results debunking false claims as early as possible to counteract public disinformation. However, this needs to be done carefully to avoid a scenario of inconsistent messaging in case the results need to be reviewed later.
- *EMLNDG (emotional nudging):* EMLNDG is referred to herein as the alternative to a factual explanation that balances facts with emotional evocation to nudge the online user to a reflective state.

Table 3 shows that four cross-dimensional trends were addressed by reviewed works from all three dimensions: PSOC, INFL, CRSPOL, and EMLNDG. Of the four, PSOC had the most representation, with 32 papers. Several PSOC constructs were present in these 32 works. Confirmation bias was shown to have a positive correlation with the openness personality trait and a negative correlation with neuroticism [161]. These two personality traits are part of the big five personality model [162]. The entrenchment of beliefs is a complex construct that may have several root drivers; however, the personality or character of the believer has been identified as an important factor [163]. Attitudinal HOMY refers to personality and attitude similarities between individuals [164]. In the context of celebrities, the more a person identifies similarities between a celebrity's attitudes and overall personality and their own, the more this individual will research about that celebrity [165], leading to parasocial relationships. Previous studies have demonstrated HOMY as one of the predictors of parasocial interactions [166]. There is currently no consensus on the meaning of HOMY beyond the broad definition stating that like-minded people tend to form communities. However, some psychology researchers argue that one's personality defines one's HOMY nature [167]. Previous research correlated the spiral of silence with the cultural behaviors of individualism and collectivism [168]. Emotion research theory considers culture to be one of the three driving influences of how people perceive and act on emotions [169]. Therefore, it is plausible, albeit still not confirmed, that cultural differences may offer a correlation between the spiral of silence and emotion. Research is evolving toward an irrefutable connection between these psychological constructs and individual personalities. However, the correlations should not be ignored.

The second ubiquitous trend with the most representation in the review is INFL, with a total 19. Trust in the information source is the foundation for the INFL's driving force in SNSs, its development ranging from parasocial relationships in the case of celebrity INFLs through confirmation biases and other intrinsic individual tendencies. A possible conclusion is that trust in INFLs has its roots in psychological constructs. Therefore, the INFL trend may be considered a corollary of the PSOC trend in our context. Psychology research has also shown a direct relationship between psychology, personality, and emotions [170–172]. We argue that this close relationship between personality and emotion may be why EMLNDG also appears in all three dimensions of this work, addressed by 14 papers.

CRSPOL is the last cross-dimensional trend that covers all three domains, albeit with a total number of papers much lower than the other three trends. INFLs have a strong effect in driving polarization, which emerges in good part due to psychological or emotional reasons or a combination of both. It is expected that a topic chiefly influenced by the three main drivers of belief change would also be present in all three review domains.

There is a strong relationship between emotion and sentiment, as sentiment can be construed as a thought, an opinion held by the person based on a feeling. In general terms, sentiment is the effect of emotion [173]. Since emotion plays

Table 3: Reviewed papers addressing each cross-dimensional trend.

	SUBCNS	FACBAL	INFL	SENTANL	PSOC	CRSPOL	ERLPUB	EMLNDG	Total
Domain specific	0	3	8	2	16	5	1	8	43
Polarization	5	0	4	0	11	3	0	4	27
Opinion dynamics	4	2	7	1	5	1	1	2	23
Total	9	5	19	3	32	9	2	14	93

such a pivotal role in belief change, it becomes natural that SENTANL emerged as a cross-domain trend and the preferred method for evaluating online belief change.

Perhaps contrary to intuition, purely factual explanations are not the most efficient in changing online forged beliefs. EMLNDG's importance to belief change drives the corollary cross-dimensional trend of FACBAL. FACBAL focuses on balancing facts and emotional arousal in explanations. The ERLPUB cross-dimensional trend is a natural consequence of the potential breach of trust between public opinion and officials who publish erroneous early communications and are forced to review the message later. The evolution of belief change from a fragmented network through the formation of subnetworks of consensus that eventually lead to polarization is the central topic of the SUBCNS cross-dimensional trend.

In summary, we showed that the cross-dimensional trends present in all three dimensions of our work are driven either by PSOC constructs, emotion, or a combination of the two. We showed that the other cross-dimensional trends presented have roots in these two drivers. Therefore, we argue that PSOC constructs and emotions are the two main drivers of online belief change. The following section will present answers to each of the proposed research questions.

6.1. Research Questions Answered. This section provides answers to research questions that emerged from the reviewed works.

6.1.1. RQ1—What Are the Main Drivers of Online Belief Change? The discussion in Section 6 presented the cross-dimensional trends, and Table 3 shows the breakdown of the number of papers that addressed each one of the trends. A numerical analysis of Table 3 indicates PSOC constructs to be the top ubiquitous trend, addressed by 34.4% of all papers. INFLs were the second most addressed trend by 20.4% of all papers. We did argue, however, that trust is a psychological construct, and it is at the center of the INFL drive for belief change. This argument suggests that both trends can be combined, leading to over half, or 54.4%, of all reviewed papers to have focused on PSOC constructs for belief change. Within the context of each dimension, combining the two trends resulted in 55.8% of the domain-specific works, 55.5% of the polarization works, and 52.1% of the opinion dynamics works. This shows an equivalent balance of relevance within each of the domains. EMLNDG accounts for a total of 15.1% of all reviewed works. This trend is the distant next highest trend, but it is much more prevalent than CRSPOL, the last ubiquitous trend, which appears in just 9.6% of all papers. We argued, however, that CRSPOL, as well as the other identified cross-dimensional trends, was corollary to the two main ones. This numerical analysis indicates that PSOC constructs and emotional arousal are arguably the two main drivers of online belief change.

Psychology research has also shown a strong correlation between individuality, personality, and emotions. Tellegen [174] has proposed that even though environmental changes may influence affective responses, a full appreciation of individual differences in emotional response could only be performed if personalities and how they influence affect are considered. The authors of [175] performed a user study and concluded that personality is an essential determinant of an individual's emotional response. Moreover, in [176], a user study shows that individuals who present with high negative affectivity are generally more introspective, a personality trait, and are more likely to experience discomfort at all times, even in the absence of stress. This shows that individuals perceive emotions differently.

Personality and individuality have been treated as synonyms by various English-language dictionaries. Personality has been defined as "the incarnation of individuality" [177]. The strong correlation between personality and emotion suggests them to be individual characteristics. Therefore, we argue that individuality is the most critical driver of online belief change, materialized through psychological traits and emotions. This result partially validates this work's central hypothesis that personalized explanations are more efficient in reducing fake news spread.

6.1.2. RQ2—How Are Current Opinion Models Being Used to Capture Belief Changes? The current opinion models used in the reviewed works that either performed simulations or used real datasets to perform experiences yielded important conclusions in capturing belief change. Interestingly, the opinion dynamics dimension works contributed to all eight cross-dimensional trends in Table 3. It is important to note how these works help to model the evolution of opinion dynamics, starting from regular social networks into multiple subnetworks of consensus and ultimately into polarization.

The most popular approach for capturing belief change is using NLP SENTANL models in social media posts [178]. The overarching concept is to perform a sentiment temporal analysis [179] of posts before and after an event with the potential to drive belief change to verify sentiment change over a specific subject. In the context of our research, a given fake claim is the subject, and the provided explanation is the event of interest. A secondary approach that has been gaining momentum is the temporal analysis of patterns

of emotions associated with social media posts [180]. This approach focuses on performing a lexicon-based analysis measuring valence, arousal, and dominance of social media posts using the VAD Lexicon [181]. The works reviewed also showed how INFLs help drive the evolution of opinion in social networks. It was demonstrated that INFLs are very important in shaping the beliefs of the subnetworks of consensus. Moreover, it was also shown how INFLs who manipulate information for some personal gain seem to have an even greater driving force in the creation of polarized networks. This is critical information for creating explanations that can efficiently change online users' preconceived beliefs. Since INFLs are key to forging opinions, they can potentially be critical agents to start a snowball effect of belief change toward a well-balanced factual explanation debunking fake news. Therefore, the models that identify social network INFLs are also important to belief change. More specifically, the models that find context-based INFLs that polarize these subnetworks [182] can be used to help focus the application of the explanation on these INFLs, followed by the application of temporal SENTANL models to verify whether the explanation was effective in changing their beliefs.

6.1.3. RQ3—What Role Does Polarization Play in Changing Online Users' Preconceived Beliefs? The highest driving forces of belief change were also central to creating or expanding polarized online scenarios. Psychological traits and INFL trends account for 55.5% of all reviewed works in this dimension. These papers confirmed that HOMY, confirmation bias, and trust in the information source are especially influential in belief change within polarized networks. Moreover, the SUBCNS trend, covered by 18.5% of the reviewed works in this dimension, presented an important characteristic of belief change toward polarization. It was concluded that a condition of polarization can be very easily created in SNSs. The evolutionary process of polarized network formation starts in opinion-segregated networks, advances to multiple subnetworks of consensus, and settles in many polarized networks attracted by and formed around INFL agents. It was seen that once polarization is established, entrenchment and backfire effects are typical psychoemotional responses by polarized individuals to factual explanations that contradict their preconceived beliefs. These individuals become somewhat immune to fact-based correcting information and, therefore, much more resistant to belief change. Two important conclusions can be drawn. Firstly, it shows the importance of balancing factual explanations with enough emotional content to gently nudge these individuals into reflective states to work around entrenchment situations. Secondly, changing an INFL's belief from within a polarized community may trigger a belief change snowball effect. This motivates the hypothesis that a belief-changing approach could combine the two concepts: adding emotional content to the explanation to reduce entrenchment, focus, and personalize these explanations on the INFLs of a polarized network.

6.1.4. RQ4—What Alternative Approaches to Offering Fact-Checked Explanations Have Been Pursued by the Literature in an Attempt to Change Preconceived Online Beliefs? The literature does not seem to offer many alternatives to fact-checked explanations. Some works evaluate whether fact-based explanations are efficient in changing beliefs; however, they have not attempted to apply alternate methods. Even though we could not find a direct answer to this RQ, it motivated the emergence of an interesting conclusion. The evaluation of efficiency in changing online users' beliefs by a purely factual explanation was performed in different domains: images, audio, and text messages. The studies come to the conclusion that factual explanations are inefficient in changing preconceived beliefs in all presentation domains and that a gentler approach should be investigated.

6.1.5. RQ5—How Can the Identified Belief-Changing Approaches Be Generalized to Multiple Belief Domains? The answer to RQ4 showed the inefficiency of purely factual explanations across all presentation domains. The same conclusion was reached by works looking at different subject domains. This aligns with our hypothesis that explanation personalization can be the nudge to drive users toward that deliberate thinking-reasoning process. The original goal of this RQ was to find out if the identified belief-changing approaches can be generally applied to multiple fake news domains. As stated, the studies did not specifically reveal alternate methodologies; however, the fact that studies in many different domains recommended the concept of nudging people into reflective states indicates that an approach of balanced explanations as previously stated may be efficient across domains.

6.2. Grand Challenges and Future Work. This section presents the grand challenges that emerged from the identification of the cross-dimension trends.

6.2.1. Psychology Research Intersection. This work argues that individuality is the chief driver of online belief change through its exteriorization as personality traits and individual emotional responses. Several psychological constructs were presented as being of influence in opinion formation. It would be important to advance online belief change and fake news explanation research to have a more mature foundation of psychology research showing solid relationships between given personality types, emotional responses, and the psychological constructs identified as important for online belief change. Even though the field shows meaningful correlations that should not be dismissed, research in this area is still evolving.

As an example, we have shown that explanations that are balanced between factual and emotional content carry a higher potential to avoid preconceived beliefs entrenchment. However, it is essential to consider that the same content may elicit different emotions in different explanation recipients. It is plausible that a given explanation that is expected to elicit a positive emotion to nudge the recipient into a reflective state may backfire and generate a counterproductive one that may drive entrenchment. A deeper understanding of how people of different personalities react to emotions could provide more qualified information to be used in a deeper personalization of the content to maximize

positive emotions. Without it, the results may be negatively biased if an imbalance of personalities is involved in the method evaluation. With these limitations in mind, future research investigating this hypothesis should include extensive demographic and cognitive preference data to characterize the study participants as much as possible. This approach may offer opportunities for cross-correlation of emotional responses with individual attributes that may shed light on potentially unexpected or contrary results.

6.2.2. INFL Detection. This work suggested that targeting well-balanced explanations on INFL agents may cause a positive snowball effect to break polarization. This requires identifying these INFLs in what may be a highly segregated network. Research in INFL identification is in its very early stages. Some works focus on this task; however, they currently propose approaches for specific domains, such as marketing [140] and health [183]. Therefore, identifying INFLs on segregated online community networks is an open research area.

6.2.3. SENTANL Versus Emotion Recognition. SENTANL was shown to be the preferred method for belief change detection. Even though SENTANL and emotion recognition are sometimes used interchangeably, they are, in fact, very different. SENTANL identifies the polarity of the person's attitude toward a given subject as positive, neutral, or negative. On the other hand, emotion recognition is the task of classifying feelings using an emotion model according to the psychology of emotions theory. This is a much more challenging goal and constitutes an entire subfield of affective computing [184]. Emotion detection could be applied automatically to detect the emotions a given explanation elicits in a given user to provide feedback for improving the explanation generation process. SENTANL can be applied to verify whether a given explanation changed the belief and then cross-referenced to the emotion evoked by the explanation for a deeper insight into the belief-changing process.

6.2.4. Polarization Prevention. It was shown in this work that initially fragmented networks evolve into subnetworks of consensus and then into polarized networks. Changing individuals' beliefs in a fragmented network may be easier than in polarized networks. For this to happen, the belief-changing system should attempt to prevent the network from becoming polarized by preempting the network's natural evolution. A preempting belief-changing system would be required to include what is known as real-time fake news detection systems. Real-time methods are receiving the community's attention [185]. They aim to identify potentially fake news at a speed compatible with the typical fake news spread speed. However, this research field is also in its early stages, offering several opportunities.

6.2.5. Well-Balanced Explanations. This work verified that gentle nudging is more efficient than cold factual counterarguments to change online beliefs. The field seems primed for explanations other than fact-heavy text. Figure 1 proposes creative explanations as a potential alternative to fact-checked text. Creative explanations can be defined as involving some level of the creative process to generate an output that can be considered creative. Creativity is immediately connected to art. Art arguably sparks experiences that simultaneously engage many aspects of an individual's mental life, including emotions [186]. Art in this context can be expanded to its multiple domains, such as poetry, music, painting, and others. Humor has been shown to include patterns of intercorrelations with several measures of creativity [187] and is a vehicle for emotional arousal [188]. In high-level terms, as long as the explanation is anchored in facts to avoid the risk of misinforming the reader, any explanation that invokes some emotional reaction could be a valid candidate to be investigated.

The applicability of the hypothesis that well-balanced explanations are more effective than purely factual ones needs to be verified in different nonpolarized and polarized subjects. The hypothesis is for beliefs surrounding nonpolarized subjects to be less challenging to change. Furthermore, polarization was shown to have a correlation between multiple subjects. Politics was shown to be a centralizing polarization topic. Therefore, it is expected to be much more challenging for a polarized individual to change their political beliefs than their opinion about another polarized subject of less centralizing power.

7. Conclusion

This work presented a systematic literature review of online belief change from the perspective of three dimensions: domain-specific user studies, polarization, and opinion dynamics. We showed evidence that PSOC constructs and emotional arousal are the two main drivers of online belief change. It was presented that this finding is in line with psychology research and that due to the close relationship of individuality with psychological constructs and emotion, individuality is arguably the single most influential force in online belief change. This finding validates the main hypothesis of this work, which states that personalization of fake news explanations is a needed improvement for fake news systems. It was also shown that all the identified cross-domain trends are rooted in individuality, demonstrating the importance of personalization to changing preconceived beliefs in fake news. Chiefly, the conclusion was that well-balanced explanations between facts and emotionally evoking content that can nudge people into a reflective state are the best candidates for the task. We also presented reasons why these types of explanations may be successful across multiple fake news domains.

Polarization was confirmed to be a strong adverse driver of belief change. We have shown alignment between polarization tendencies and individuality. Entrenchment and backfire effect are two constructs that work against belief change and become especially strong as a reaction to purely factual explanations contradicting preconceived beliefs. Polarization has been shown to be cross-correlated, with politics arguably being the central polarization pole. Polarized individuals with a specific political ideology also tend to be polarized on other subjects such as climate change,

immigration, policy, COVID-19 response, minority rights, and other sensitive topics. Trust is one of the strongest psychological drivers of polarization. For this reason, INFLs become critical agents of polarization, especially the ones who purposely manipulate information for some form of personal gain. Furthermore, it was concluded that segregated social networks of opinion evolve through the formation of subnetworks of consensus and ultimately to polarized online social groups. These two findings can potentially be used in favor of fake news debunking systems by delivering well-balanced explanations to polarized network INFLs. Since they are driving forces of opinion formation, changing the preconceived beliefs of just a few of these agents may create a favorable snowball effect in the entire social network toward consensus against fake news.

Data Availability Statement

The findings supporting this systematic review are from previously reported studies and datasets, which have been cited. The processed data are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Funding

This research was supported by Fundação para a Ciência e Tecnologia (FCT) through (1) INESC-ID Multiannual Funding with reference UIDB/50021/2020 (doi:10.54499/UIDB/50021/2020); (2) Grant 2022.09212.PTDC (XAVIER Project) and project UIDB/50021/2020 (doi:10.54499/UIDB/50021/2020), under the auspices of the UNESCO Chair on AI&VR of the University of Lisbon; and (3) CHIST-ERA within the CIMPLE project (CHIST-ERA-19-XAI-003) which corresponds to the FCT reference CHIST-ERA/0001/2019.

References

[1] Y. M. Rocha, G. A. de Moura, G. A. Desidério, C. H. de Oliveira, F. D. Lourenço, and L. D. de Figueiredo Nicolete, "The impact of fake news on social media and its influence on health during the COVID-19 pandemic: a systematic review," *Journal of Public Health*, vol. 31, no. 7, pp. 1007–1016, 2023.

[2] S. A. García, G. G. García, M. S. Prieto, A. J. M. Guerrero, and C. R. Jiménez, "The impact of term fake news on the scientific community. Scientific performance and mapping in web of science," *Social Sciences*, vol. 9, no. 5, p. 73, 2020.

[3] K. Sipitanos, "Raising awareness against fake news to protect democracy: the myth of islamophobia in Trump's speech," *Social Semiotics*, vol. 33, no. 4, pp. 714–730, 2023.

[4] A. Mitchell, J. Gottfried, G. Stocking, M. Walker, and S. Fedeli, "Many Americans say made-up news is a critical problem that needs to be fixed," *Pew Research Center*, vol. 5, p. 2019, 2019.

[5] Y. Mirsky and W. Lee, "The creation and detection of deepfakes: a survey," *ACM Computing Surveys (CSUR)*, vol. 54, no. 1, pp. 1–41, 2021.

[6] A. Satariano and P. Mozur, "The people onscreen are fake. The disinformation is real," 2023, https://www.nytimes.com/2023/02/07/technology/artificial-intelligence-training-deepfake.html.

[7] P. Verma, "They thought loved ones were calling for help. It was an AI scam," 2023, https://www.washingtonpost.com/technology/2023/03/05/ai-voice-scam/.

[8] K. Mishima and H. Yamana, "A survey on explainable fake news detection," *IEICE Transactions on Information and Systems*, vol. E105.D, no. 7, pp. 1249–1257, 2022.

[9] A. Raj and M. P. Goswami, "Is fake news spreading more rapidly than COVID-19 in India," *Journal of Content, Community and Communication*, vol. 11, no. 10, pp. 208–220, 2020.

[10] R. Denaux, M. Mensio, J. M. Gomez-Perez, and H. Alani, "Weaving a semantic web of credibility reviews for explainable misinformation detection (extended abstract)," in *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, p. 1, Montreal, Canada, August 2021.

[11] F. Yang, S. K. Pentyala, S. Mohseni et al., "Xfake: explainable fake news detector with visualizations," in *The World Wide Web Conference*, pp. 3600–3604, New York, NY, USA, May 2019.

[12] K. Shu, L. Cui, S. Wang, D. Lee, and H. Liu, "defend: explainable fake news detection," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 395–405, New York, NY, 2019.

[13] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, and G. Z. Yang, "Xai—explainable artificial intelligence," *Science Robotics*, vol. 4, no. 37, 2019.

[14] A. Adadi and M. Berrada, "Peeking inside the black-box: a survey on explainable artificial intelligence (xai)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.

[15] X. Zhang and A. A. Ghorbani, "An overview of online fake news: characterization, detection, and discussion," *Information Processing and Management*, vol. 57, no. 2, article 102025, 2020.

[16] X. Chen, P. Tsaparas, J. Lijffijt, and T. De Bie, "Opinion dynamics with backfire effect and biased assimilation," *PLoS One*, vol. 16, no. 9, article e0256922, 2021.

[17] P. Atanasova, J. G. Simonsen, C. Lioma, and I. Augenstein, "Generating fact checking explanations," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 7352–7364, July 2020.

[18] M. Ruffin, G. Wang, and K. Levchenko, "Explaining why fake photos are fake: Does it work?," *Proceedings of the ACM on Human-Computer Interaction*, vol. 7, no. GROUP, pp. 1–22, 2023.

[19] W. Zhang and B. Y. Lim, "Towards relatable explainable AI with the perceptual process," in *CHI Conference on Human Factors in Computing Systems*, pp. 1–24, New York, NY, April 2022.

[20] N. Bryan-Kinns, C. Ford, A. Chamberlain et al., "Explainable AI for the arts: XAIxArts," in *Creativity and Cognition*, pp. 1–7, New York, NY, USA, June 2023.

[21] A. Caraban, E. Karapanos, D. Gonçalves, and P. Campos, "23 ways to nudge: a review of technology-mediated nudging in human-computer interaction," in *Proceedings of the 2019*

CHI Conference on Human Factors in Computing Systems, pp. 1–15, New York, NY, USA, May 2019.

[22] F. Altoe and H. S. Pinto, "Towards a personalized online fake news taxonomy," in *Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization*, pp. 96–105, New York, NY, USA, June 2023.

[23] S. Colton G. A. Wiggins et al., "Computational creativity: the final frontier?," in *Ecai*, vol. 12, pp. 21–26, Montpelier, Automated Computer Machinery, New York, NY, 2012.

[24] A. Jordanous, "Four pppperspectives on computational creativity in theory and in practice," *Connection Science*, vol. 28, no. 2, pp. 194–216, 2016.

[25] D. W. MacKinnon, "Creativity: a multi-faceted phenomenon," in *Creativity: A Discussion at the Nobel Conference*, J. D. Roslansky, Ed., pp. 17–32, North-Holland, Amsterdan, 1970.

[26] M. Rhodes, "An analysis of creativity," *The Phi Delta Kappan*, vol. 42, no. 7, pp. 305–310, 1961.

[27] C. Lamb, D. G. Brown, and C. L. Clarke, "Evaluating computational creativity: an interdisciplinary tutorial," *ACM Computing Surveys (CSUR)*, vol. 51, pp. 1–34, 2018.

[28] M. Riedl, "Weird AI Yankovic: generating parody lyrics," 2020, https://arxiv.org/abs/2009.12240.

[29] S. Sharma, S. Agarwal, T. Suresh, P. Nakov, M. S. Akhtar, and T. Chakraborty, "What do you meme? Generating explanations for visual semantic role labelling in memes," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 8, pp. 9763–9771, 2023.

[30] N. Köbis and L. D. Mossink, "Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry," *Computers in Human Behavior*, vol. 114, article 106553, 2021.

[31] J. Toplyn, "Witscript: a system for generating improvised jokes in a conversation," 2023, https://arxiv.org/abs/2302.02008.

[32] Y. Chang, X. Wang, J. Wang et al., "A survey on evaluation of large language models," *ACM Transactions on Intelligent Systems and Technology*, vol. 15, no. 3, pp. 1–45, 2024.

[33] Y. Wang, W. Zhong, L. Li et al., "Aligning large language models with human: a survey," 2023, https://arxiv.org/abs/2307.12966.

[34] P. Quintas and H. S. Pinto, *Report on the state of the art on creative xai*, vol. 5.1, Tech. rep., CIMPLE project deliverable, 2022.

[35] H. Noorazar, K. R. Vixie, A. Talebanpour, and Y. Hu, "From classical to modern opinion dynamics," *International Journal of Modern Physics C*, vol. 31, no. 7, article 2050101, 2020.

[36] H. Noorazar, "Recent advances in opinion propagation dynamics: a 2020 survey," *The European Physical Journal Plus*, vol. 135, p. 521, 2020.

[37] R. Ureña, G. Kou, Y. Dong, F. Chiclana, and E. Herrera-Viedma, "A review on trust propagation and opinion dynamics in social networks and group decision making frameworks," *Information Sciences*, vol. 478, pp. 461–475, 2019.

[38] L. Mastroeni, P. Vellucci, and M. Naldi, "Agent-based models for opinion formation: a bibliographic survey," *IEEE Access*, vol. 7, pp. 58836–58848, 2019.

[39] O. Abid, S. Jamoussi, and Y. B. Ayed, "Deterministic models for opinion formation through communication: a survey," *Online Social Networks and Media*, vol. 6, pp. 1–17, 2018.

[40] Y. Wang, M. McKee, A. Torbica, and D. Stuckler, "Systematic literature review on the spread of health-related misinformation on social media," *Social Science and Medicine*, vol. 240, article 112552, 2019.

[41] K. Tsamakis, D. Tsiptsios, B. Stubbs et al., "Summarising data and factors associated with COVID19 related conspiracy theories in the first year of the pandemic: a systematic review and narrative synthesis," *BMC Psychology*, vol. 10, no. 1, p. 244, 2022.

[42] L. Iandoli, S. Primario, and G. Zollo, "The impact of group polarization on the quality of online debate in social media: a systematic literature review," *Technological Forecasting and Social Change*, vol. 170, article 120924, 2021.

[43] J. Khalid, A. Abbas, R. Akbar et al., "Significance of electronic word of mouth (e-wom) in opinion formation," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 2, 2020.

[44] H. Xie, M. Zhong, Y. Li, and J. C. S. Lui, "Understanding persuasion cascades in online product rating systems: modeling, analysis, and inference," *ACM Transactions on Knowledge Discovery from Data*, vol. 15, no. 3, pp. 1–29, 2021.

[45] A. Ardèvol-Abreu, H. G. de Zúñiga, and E. Gámez, "The influence of conspiracy beliefs on conventional and unconventional forms of political participation: the mediating role of political efficacy," *British Journal of Social Psychology*, vol. 59, no. 2, pp. 549–569, 2020.

[46] C. Ruess, C. P. Hoffmann, S. Boulianne, and K. Heger, "Online political participation: the evolution of a concept," *Information, Communication and Society*, vol. 26, no. 8, pp. 1495–1512, 2023.

[47] M. Galesic, H. Olsson, J. Dalege, T. van der Does, and D. L. Stein, "Integrating social and cognitive aspects of belief dynamics: towards a unifying framework," *Journal of the Royal Society Interface*, vol. 18, no. 176, article 20200857, 2021.

[48] S. D. Arora, G. P. Singh, A. Chakraborty, and M. Maity, "Polarization and social media: a systematic review and research agenda," *Technological Forecasting and Social Change*, vol. 183, article 121942, 2022.

[49] A.-M. Bliuc, A. Bouguettaya, and K. D. Felise, "Online intergroup polarization across political fault lines: an integrative review," *Frontiers in Psychology*, vol. 12, article 641215, 2021.

[50] G. Yang, *Multidisciplinary Studies in Knowledge and Systems Science*, IGI Global, Beijing, China, 2013.

[51] N. Van Eck and L. Waltman, "Software survey: Vosviewer, a computer program for bibliometric mapping," *Scientometrics*, vol. 84, no. 2, pp. 523–538, 2010.

[52] R. A. Varanasi, J. Pal, and A. Vashistha, "Accost, accede, or amplify: attitudes towards COVID-19 misinformation on WhatsApp in India," in *CHI Conference on Human Factors in Computing Systems*, pp. 1–17, New York, NY, April 2022.

[53] T. J. Padamsee, R. M. Bond, G. N. Dixon et al., "Changes in COVID-19 vaccine hesitancy among Black and White individuals in the US," *JAMA Network Open*, vol. 5, no. 1, article e2144470, 2022.

[54] T. Do, D. Nguyen, A. Le et al., "Understanding public opinion on using hydroxychloroquine for COVID-19 treatment via social media," 2022, http://arxiv.org/abs/2201.00237.

[55] T. Chen, L. Peng, J. Yang, and G. Cong, "Modeling, simulation, and case analysis of COVID19 over network public opinion formation with individual internal factors and external

information characteristics," *Concurrency and Computation: Practice and Experience*, vol. 33, no. 17, article e6201, 2021.

[56] R. Zhang, Y. Shen, C. Lin, and H. Li, "Evolution of public opinion on COVID-19 based on microblog visualization," in *2022 the 5th International Conference on Information Science and Systems*, pp. 142–148, New York, NY, USA, August 2022.

[57] G. Klaus, A. Ernst, and L. Oswald, "Psychological factors influencing laypersons' acceptance of climate engineering, climate change mitigation and business as usual scenarios," *Technology in Society*, vol. 60, article 101222, 2020.

[58] T. J. B. Cann, I. S. Weaver, and H. T. P. Williams, "Ideological biases in social sharing of online information about climate change," *PLoS One*, vol. 16, no. 4, article e0250656, 2021.

[59] A. Shehata, J. Johansson, B. Johansson, and K. Andersen, "Climate change frame acceptance and resistance: extreme weather, consonant news, and personal media orientations," *Mass Communication and Society*, vol. 25, no. 1, pp. 51–76, 2022.

[60] P. P. Sun, "Understanding EFL university teachers' synchronous online teaching belief change," *Language Teaching Research*, vol. 1, article 136216882210938, 2022.

[61] R. C. Seoane and J. E. Jiménez, "Effectiveness of online and blended delivery methods on preservice teachers' knowledge and beliefs for writing instruction," *Journal of Education for Teaching*, vol. 48, no. 2, pp. 178–196, 2022.

[62] A. E. Flanigan, M. S. Peteranetz, D. F. Shell, and L.-K. Soh, "Shifting beliefs in computer science: change in CS student mindsets," *ACM Transactions on Computing Education*, vol. 22, no. 2, pp. 1–24, 2022.

[63] A. Best, "Primary school teachers' beliefs on computer science as a discipline and as a school subject," in *Proceedings of the 15th Workshop in Primary and Secondary Computing Education, WiPSCE 2020*, vol. 1, p. 1, Association for Computing Machinery, 2020.

[64] L. L. Gelauff and A. Goel, "Opinion change or differential turnout: Austin's budget feedback exercise and the police department," in *Equity and Access in Algorithms, Mechanisms, and Optimization*, p. 1, New York, NY, USA, October 2022.

[65] Y. Tsfati, J. Cohen, S. Dvir-Gvirsman, K. Tsuriel, I. Waismel-Manor, and R. L. Holbert, "Political para-social relationship as a predictor of voting preferences in the Israeli 2019 elections," *Communication Research*, vol. 49, no. 8, pp. 1118–1147, 2022.

[66] J.-P. Fränken, N. Theodoropoulos, A. Moore, and N. Bramley, "Belief revision in a micro-social network: modeling sensitivity to statistical dependencies in social learning," *CogSci*, vol. 1, pp. 1255–1261, 2020.

[67] J. Massachs, C. Monti, G. D. F. De Francisci Morales, and F. Bonchi, "Roots of Trumpism: homophily and social feedback in Donald Trump support on Reddit," in *12th ACM Conference on Web Science, Association for Computing Machinery*, pp. 49–58, New York, NY, USA, July 2020.

[68] R. J. Garcia, E. V. Shaw, and N. Scurich, "Normative and informational influence in group decision making: effects of majority opinion and anonymity on voting behavior and belief change," *Group Dynamics: Theory, Research, and Practice*, vol. 25, no. 4, pp. 319–333, 2021.

[69] F. Rigoli, "Opinions about immigration, patriotism, and welfare policies during the coronavirus emergency: the role of political orientation and anxiety," *The Social Science Journal*, vol. 57, pp. 1–10, 2020.

[70] C. Rafkin, A. Shreekumar, and P.-L. Vautrey, "When guidance changes: government stances and public beliefs," *Journal of Public Economics*, vol. 196, article 104319, 2021.

[71] M. M. Bowers and C. T. Whitley, "What drives support for transgender rights? Assessing the effects of biological attribution on U.S. public opinion of transgender rights," *Sex Roles*, vol. 83, no. 7-8, pp. 399–411, 2020.

[72] M. O. Ukonu, L. I. Anorue, U. Ololo, and H. M. Olawoyin, "Climate of conformism: social media users' opinion on homosexuality in Nigeria," *SAGE Open*, vol. 11, no. 3, article 215824402110407, 2021.

[73] A. Freedle and S. Kashubeck-West, "Core belief challenge, rumination, and posttraumatic growth in women following pregnancy loss," *Psychological Trauma: Theory, Research, Practice, and Policy*, vol. 13, no. 2, pp. 157–164, 2021.

[74] S. Lee, L. Atkinson, and Y. H. Sung, "Online bandwagon effects: quantitative versus qualitative cues in online comments sections," *New Media and Society*, vol. 24, no. 3, pp. 580–599, 2022.

[75] W. S. Rossi, J. W. Polderman, and P. Frasca, "The closed loop between opinion formation and personalized recommendations," *IEEE Transactions on Control of Network Systems*, vol. 9, no. 3, pp. 1092–1103, 2022.

[76] D. Barman and O. Conlan, "Exploring the links between personality traits and susceptibility to disinformation," in *Proceedings of the 32st ACM Conference on Hypertext and Social Media*, pp. 291–294, New York, NY, August 2021.

[77] A. H. Y. Chan, R. Horne, H. Lycett et al., "Changing patient and public beliefs about antimicrobials and antimicrobial resistance (AMR) using a brief digital intervention," *Frontiers in Pharmacology*, vol. 12, article 608971, 2021.

[78] I.'. I. L. D. Pinto, N. Rungratsameetaweemana, K. Flaherty et al., "Intermittent brain network reconfigurations and the resistance to social media influence," *Network Neuroscience*, vol. 6, no. 3, pp. 870–896, 2022.

[79] M. Vlasceanu, M. J. Morais, and A. Coman, "Network structure impacts the synchronization of collective beliefs," *Journal of Cognition and Culture*, vol. 21, no. 5, pp. 431–448, 2021.

[80] V. Grimm and F. Mengel, "Experiments on belief formation in networks," *Journal of the European Economic Association*, vol. 18, no. 1, pp. 49–82, 2020.

[81] M. Xu, Z. Luo, R. Liu, B. Wang, and H. Xu, "One-sided versus two-sided: a novel opinion dynamics information-type education-based Hegselmann–Krause model," in *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, p. 339-334, Melbourne, Australia, 2021.

[82] A. Mansouri and F. Taghiyareh, "Effect of segregation on opinion formation in scale-free social networks: an agent-based approach," *International Journal of Engineering*, vol. 34, no. 1, pp. 66–74, 2021.

[83] Y. Peng, Y. Zhao, and J. Hu, "On the role of community structure in evolution of opinion formation: a new bounded confidence opinion dynamics," *Information Sciences*, vol. 621, pp. 672–690, 2023.

[84] H. Takesue, "A noisy opinion formation model with two opposing mass media," 2020, https://arxiv.org/abs/2011.13813.

[85] L. Burbach, P. Halbach, M. Ziefle, and A. Calero Valdez, "Opinion formation on the internet: the influence of

personality, network structure, and content on sharing messages online," *Frontiers in Artificial Intelligence*, vol. 3, p. 45, 2020.

[86] T. Chen, X. Yin, J. Yang, G. Cong, and G. Li, "Modeling multi-dimensional public opinion process based on complex network dynamics model in the context of derived topics," *Axioms*, vol. 10, no. 4, p. 270, 2021.

[87] M. Bashari and M.-R. Akbarzadeh-T, "Theoretical development of a probabilistic fuzzy model for opinion formation in social networks," *Fuzzy Sets and Systems*, vol. 454, pp. 125–148, 2023.

[88] Z. Wu, Q. Zhou, Y. Dong, J. Xu, A. H. Altalhi, and F. Herrera, "Mixed opinion dynamics based on DeGroot model and Hegselmann–Krause model in social networks," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 1, pp. 296–308, 2023.

[89] Z. Zhang, S. Al-Abri, and F. Zhang, "Opinion dynamics on the sphere for stable consensus and stable bipartite dissensus," *IFAC-PapersOnLine*, vol. 55, no. 13, pp. 288–293, 2022.

[90] M. Bashari and M.-R. Akbarzadeh-T, "Controlling opinions in Deffuant model by reconfiguring the network topology," *Physica A: Statistical Mechanics and its Applications*, vol. 544, article 123462, 2020.

[91] I. V. Kozitsin, "A general framework to link theory and empirics in opinion formation models," *Scientific Reports*, vol. 12, no. 1, p. 5543, 2022.

[92] J. Wei, Y. Jia, Y. Zhang, H. Zhu, and W. Huang, "The public opinion evolution under group interaction in different information features," *Complexity*, vol. 2022, Article ID 1016692, 15 pages, 2022.

[93] A. Mansouri and F. Taghiyareh, "Phase transition in the social impact model of opinion formation in log-normal networks," *Journal of Information Systems and Telecommunication (JIST)*, vol. 9, no. 33, pp. 1–14, 2021.

[94] M. Liu and L. Rong, "An online multi-dimensional opinion dynamic model with misinformation diffusion in emergency events," *Journal of Information Science*, vol. 48, no. 5, pp. 640–659, 2022.

[95] Q. Zhou and Z. Wu, "Multidimensional Friedkin-Johnsen model with increasing stubbornness in social networks," *Information Sciences*, vol. 600, pp. 170–188, 2022.

[96] Y. Luo, C. Cheng, Y. Li, and C. Yu, "Opinion formation with zealots on temporal network," *Communications in Nonlinear Science and Numerical Simulation*, vol. 98, article 105772, 2021.

[97] Y. Li, Z. Chen, and H. V. Zhao, "Robust opinion control under network perturbation," *IEEE Signal Processing Letters*, vol. 29, pp. 1649–1653, 2022.

[98] S. Cheng, C. Sun, S. Yang, M. Xu, and H. Xu, "Jumping on the bandwagon: group opinion prompts agents to reach consensus," in *2021 IEEE 23rd Int Conf on High Performance Computing & Communications; 7th Int Conf on Data Science & Systems; 19th Int Conf on Smart City; 7th Int Conf on Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys)*, pp. 399–404, Haikou, Hainan, China, December 2021.

[99] S. Gündüç, "The effect of social media on shaping individuals opinion formation," in *Complex Networks and their Applications VIII: Volume 2 Proceedings of the Eighth International Conference on Complex Networks and their Applications COMPLEX NETWORKS 2019 8*, pp. 376–386, Springer, Pensilvania, PA, 2019.

[100] Z. Li, X. Tang, and Z. Hong, "Opinion dynamics induced by agents with particular goal," *Journal of Systems Science and Complexity*, vol. 35, no. 6, pp. 2319–2335, 2022.

[101] I. V. Kozitsin, "Formal models of opinion formation and their application to real data: evidence from online social networks," *The Journal of Mathematical Sociology*, vol. 46, no. 2, pp. 120–147, 2022.

[102] I. V. Pasquetto, E. Jahani, S. Atreja, and M. Baum, "Social debunking of misinformation on WhatsApp: the case for strong and in-group ties," *Proceedings of the ACM on Human-Computer Interaction*, vol. 6, no. CSCW1, pp. 1–35, 2022.

[103] Y. Tang, J. Liu, and W. Chen, "Exchange, adopt, evolve: modeling the spreading of opinions through cognition and interaction in a social network," *Information Sciences*, vol. 551, pp. 1–22, 2021.

[104] H.-J. Geiss, F. Sakketou, and L. Flek, "Ok boomer: probing the socio-demographic divide in echo chambers," in *Proceedings of the Tenth International Workshop on Natural Language Processing for Social Media*, pp. 83–105, Seattle, Washington, July 2022.

[105] T. Ha, "Understanding of majority opinion formation in online environments through statistical analysis of news, documentary, and comedy YouTube channels," *Social Science Computer Review*, vol. 41, no. 2, pp. 353–369, 2022.

[106] C. Monti, G. De Francisci Morales, and F. Bonchi, "Learning opinion dynamics from social traces," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 764–773, New York, NY, USA, August 2020.

[107] J. Ding, M. Xu, Y. K. Tse, K. Y. Lin, and M. Zhang, "Customer opinions mining through social media: insights from sustainability fraud crisis-Volkswagen emissions scandal," *Enterprise Information Systems*, vol. 17, no. 8, article 2130012, 2023.

[108] T. Rudnyk and O. Chertov, "Method for identifying twitter accounts that have changed their opinion about politicians," in *ITS*, pp. 24–35, CEUR-WS, Kyiev, Ukraine, 2020.

[109] E. Biondi, C. Boldrini, A. Passarella, and M. Conti, "Dynamics of opinion polarization," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 9, pp. 5381–5392, 2023.

[110] W. Xu, L. Zhu, J. Guan, Z. Zhang, and Z. Zhang, "Effects of stubbornness on opinion dynamics," in *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pp. 2321–2330, New York, NY, October 2022.

[111] T. Kinoshita and M. Aida, "A spectral-based model for describing social polarization in online communities," *IEICE Transactions on Communications*, vol. E105.B, no. 10, pp. 1181–1191, 2022.

[112] J. Gaitonde, J. Kleinberg, and É. Tardos, "Polarization in geometric opinion dynamics," in *Proceedings of the 22nd ACM Conference on Economics and Computation*, pp. 499–519, New York, NY, July 2021.

[113] H. P. Maia, S. C. Ferreira, and M. L. Martins, "Adaptive network approach for emergence of societal bubbles," *Physica A: Statistical Mechanics and its Applications*, vol. 572, article 125588, 2021.

[114] H. A. Prasetya and T. Murata, "A model of opinion and propagation structure polarization in social media," *Computational Social Networks*, vol. 7, no. 1, pp. 1–35, 2020.

[115] T. Donkers and J. Ziegler, "The dual echo chamber: modeling social media polarization for interventional recommending," in *Fifteenth ACM Conference on Recommender Systems*, pp. 12–22, New York, NY, September 2021.

[116] L. Wang, C. A. Wang, and X. Yao, "Befriended to polarise? The impact of friend identity on review polarisation—A-quasi-experiment," *Information Systems Journal*, vol. 1, 2023.

[117] H. Min, J. Cao, J. Ge, and B. Liu, "A multi-agent system for fine-grained opinion dynamics analysis in online social networks," *IEEE Transactions on Computational Social Systems*, vol. 11, no. 1, pp. 815–828, 2022.

[118] C. S. R. Avuthu, M. Maleszka, and N. Van Sinh, "Interchangeability of knowledge and opinion integration strategies in collective models," in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 2196–2200, Toronto, ON, Canada, October 2020.

[119] N. Hirakura, M. Aida, and K. Kawashima, "Modeling polarization caused by empathetic and repulsive reaction in online social network," *IEICE Transactions on Communications*, vol. E105.B, no. 8, pp. 990–1001, 2022.

[120] S. Tu and S. Neumann, "A viral marketing-based model for opinion dynamics in online social networks," in *Proceedings of the ACM Web Conference 2022*, pp. 1570–1578, New York, NY, April 2022.

[121] H. Ferraz de Arruda, F. Maciel Cardoso, G. Ferraz de Arruda, A. R. Hernández, L. da Fontoura Costa, and Y. Moreno, "Modelling how social network algorithms can influence opinion polarization," *Information Sciences*, vol. 588, pp. 265–278, 2022.

[122] U. Chitra and C. Musco, "Analyzing the impact of filter bubbles on social network polarization," in *Proceedings of the 13th International Conference on Web Search and Data Mining*, pp. 115–123, New York, NY, USA, January 2020.

[123] N. Loy, M. Raviola, and A. Tosin, "Opinion polarization in social networks," *Philosophical Transactions of the Royal Society A*, vol. 380, no. 2224, article 20210158, 2022.

[124] D. A. Gubanov, I. V. Petrov, and A. G. Chkhartishvili, "Multidimensional model of opinion dynamics in social networks: polarization indices," *Automation and Remote Control*, vol. 82, no. 10, pp. 1802–1811, 2021.

[125] M. W. Macy, M. Ma, D. R. Tabin, J. Gao, and B. K. Szymanski, "Polarization and tipping points," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 118, no. 50, 2021.

[126] S. Gupta, G. Jain, and A. A. Tiwari, "Polarised social media discourse during COVID-19 pandemic: evidence from YouTube," *Behaviour & Information Technology*, vol. 42, no. 2, pp. 227–248, 2023.

[127] J. Bernacer, J. García-Manglano, E. Camina, and F. Güell, "Polarization of beliefs as a consequence of the COVID-19 pandemic: The case of spain," *PLoS One*, vol. 16, no. 7, article e0254511, 2021.

[128] S. Nair, A. Iamnitchi, and J. Skvoretz, "Promoting social conventions across polarized networks: an empirical study," in *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 349–352, New York, NY, USA, August 2019.

[129] G. De Francisci Morales, C. Monti, and M. Starnini, "No echo in the chambers of political interactions on Reddit," *Scientific Reports*, vol. 11, no. 1, p. 2818, 2021.

[130] P. V. Sheela and F. Mannering, "The effect of information on changing opinions toward autonomous vehicle adoption: an exploratory analysis," *International Journal of Sustainable Transportation*, vol. 14, no. 6, pp. 475–487, 2020.

[131] S. M. Coles and M. Saleem, "Social media expression and user predispositions: applying the differential susceptibility to media effects model to the study of issue polarization," *Social Media + Society*, vol. 7, no. 4, article 205630512110529, 2021.

[132] A. Tyagi, J. Uyheng, and K. M. Carley, "Affective polarization in online climate change discourse on twitter," in *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 443–447, The Hague, Netherlands, December 2020.

[133] A. Spitz, A. Abu-Akel, and R. West, "Interventions for softening can lead to hardening of opinions: evidence from a randomized controlled trial," in *Proceedings of the Web Conference 2021*no. p, pp. 1098–1109, New York, NY, USA, April 2021.

[134] N. Yeung, J. Lai, and J. Luo, "Face off: polarized public opinions on personal face mask usage during the COVID-19 pandemic," in *2020 IEEE International Conference on Big Data (Big Data)*, pp. 4802–4810, Atlanta, GA, USA, December 2020.

[135] S. Perrett, "A divided kingdom? Variation in polarization, sorting, and dimensional alignment among the British public, 1986–2018," *British Journal of Sociology*, vol. 72, no. 4, pp. 992–1014, 2021.

[136] C. Largeron, A. Mardale, and M. A. Rizoiu, "Linking the dynamics of user stance to the structure of online discussions," in *Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Lecture Notes in Computer Science ((LNISA,volume 12695)), Springer, Pensilvania, PA, 2021.

[137] J. J. B. Mijs, W. de Koster, and J. van der Waal, "Belief change in times of crisis: providing facts about COVID-19-induced inequalities closes the partisan divide but fuels intra-partisan polarization about inequality," *Social Science Research*, vol. 104, article 102692, 2022.

[138] G. Fischer, P. Riedesser, and A. G. Fischer, *Lehrbuch der psychotraumatologie*, Ernst Reinhardt Verlag München, München, DE, 2020.

[139] A. Boals, "Trauma in the eye of the beholder: objective and subjective definitions of trauma," *Journal of Psychotherapy Integration*, vol. 28, no. 1, pp. 77–89, 2018.

[140] P. Harrigan, T. M. Daly, K. Coussement, J. A. Lee, G. N. Soutar, and U. Evers, "Identifying influencers on social media," *International Journal of Information Management*, vol. 56, article 102246, 2021.

[141] Z. Liberman, K. D. Kinzler, and A. L. Woodward, "Origins of homophily: Infants expect people with shared preferences to affiliate," *Cognition*, vol. 212, article 104695, 2021.

[142] M. Cinelli, G. De Francisci Morales, A. Galeazzi, W. Quattrociocchi, and M. Starnini, "The echo chamber effect on social media," *Proceedings of the National Academy of Sciences*, vol. 118, no. 9, article e2023301118, 2021.

[143] R. S. Nickerson, "Confirmation bias: a ubiquitous phenomenon in many guises," *Review of General Psychology*, vol. 2, no. 2, pp. 175–220, 1998.

[144] T. T. Nguyen, P.-M. Hui, F. M. Harper, L. Terveen, and J. A. Konstan, "Exploring the filter bubble: The effect of using recommender systems on content diversity," in *Proceedings of*

the 23rd international conference on World wide web, pp. 677–686, New York, NY, USA, April 2014.

[145] D. M. Rousseau, S. B. Sitkin, R. S. Burt, and C. Camerer, "Not so different after all: a cross-discipline view of trust," *Academy of Management Review*, vol. 23, no. 3, pp. 393–404, 1998.

[146] R. F. Baumeister and M. R. Leary, "The need to belong: desire for interpersonal attachments as a fundamental human motivation," *Psychological Bulletin*, vol. 117, no. 3, pp. 497–529, 1995.

[147] G. D. Bishop and D. G. Myers, "Informational influence in group discussion," *Organizational Behavior and Human Performance*, vol. 12, no. 1, pp. 92–104, 1974.

[148] E. Noelle-Neumann, *The Spiral of Silence: Public Opinion–Our Social Skin*, University of Chicago Press, Chicago, IL, 1993.

[149] T.-Y. Wu and C. A. Lin, "Predicting the effects of eWOM and online brand messaging: source trust, bandwagon effect and innovation adoption factors," *Telematics and Informatics*, vol. 34, no. 2, pp. 470–480, 2017.

[150] K. Tapson, M. Doyle, V. Karagiannopoulos, and P. Lee, "Understanding moral injury and belief change in the experiences of police online child sex crime investigators: an interpretative phenomenological analysis," *Journal of Police and Criminal Psychology*, vol. 37, no. 3, pp. 637–649, 2022.

[151] L. G. Calhoun and R. G. Tedeschi, *Posttraumatic growth: The positive lessons of loss*, American Psychological Association, 2001.

[152] W. Xu, H. Jiang, Y. Zhou, L. Zhou, and H. Fu, "Intrusive rumination, deliberate rumination, and posttraumatic growth among adolescents after a tornado: the role of social support," *The Journal of Nervous and Mental Disease*, vol. 207, no. 3, pp. 152–156, 2019.

[153] A. M. Brandt, "Racism and research: the case of the Tuskegee syphilis study," *Hastings Center Report*, vol. 8, no. 6, pp. 21–29, 1978.

[154] S. Zielinski, "Henrietta lacks 'immortal'cells," *Smithsonian Magazine*, vol. 22, 2010.

[155] H. M. Boehme, R. J. Kaminski, and M. S. Nolan, "City-wide firearm violence spikes in Minneapolis following the murder of George Floyd: a comparative time-series analysis of three cities," *Urban Science*, vol. 6, no. 1, p. 16, 2022.

[156] N. Schwarz and G. L. Clore, "Mood as information: 20 years later," *Psychological Inquiry*, vol. 14, no. 3, pp. 296–303, 2003.

[157] R. B. Cialdini and N. J. Goldstein, "Social influence: compliance and conformity," *Annual Review of Psychology*, vol. 55, no. 1, pp. 591–621, 2004.

[158] K. H. Jamieson and J. N. Cappella, *Echo Chamber: Rush Limbaugh and the Conservative Media Establishment*, Oxford University Press, Oxford, EN, 2008.

[159] E. Pariser, *The filter bubble: what the Internet is hiding from you*, Penguin UK, London, England, 2011.

[160] A. J. Stewart, M. Mosleh, M. Diakonova, A. A. Arechar, D. G. Rand, and J. B. Plotkin, "Information gerrymandering and undemocratic decisions," *Nature*, vol. 573, no. 7772, pp. 117–121, 2019.

[161] A. Melinder, T. Brennen, M. F. Husby, and O. Vassend, "Personality, confirmation bias, and forensic interviewing performance," *Applied Cognitive Psychology*, vol. 34, no. 5, pp. 961–971, 2020.

[162] B. De Raad, *The Big Five Personality Factors: The Psycholexical Approach to Personality*, Hogrefe & Huber Publishers, Newburyport, MA, 2000.

[163] P. C. Smith, *Promoting Belief Change by Encouraging Evaluation of Prior Beliefs*, The University of Wisconsin-Milwaukee, Milwaukee, WI, 2000.

[164] C. R. Berger and R. J. Calabrese, "Some explorations in initial interaction and beyond: toward a developmental theory of interpersonal communication," *Human Communication Research*, vol. 1, no. 2, pp. 99–112, 1975.

[165] J.-J. Igartua and I. Barrios, "Changing real-world beliefs with controversial movies: processes and mechanisms of narrative persuasion," *Journal of Communication*, vol. 62, no. 3, pp. 514–531, 2012.

[166] D. Bao, D. Dong, and X. Meng, "Parasocial interaction between browser and poster in virtual communities: an empirical study on dianping. com," *Chinese Journal of Management*, vol. 8, no. 7, pp. 1010–1020, 2011.

[167] R. S. Solomon, P. Y. K. L. Srinivas, A. Das, B. Gamback, and T. Chakraborty, "Understanding the psycho-sociological facets of homophily in social network communities," *IEEE Computational Intelligence Magazine*, vol. 14, no. 2, pp. 28–40, 2019.

[168] H. Huang, "A cross-cultural test of the spiral of silence," *International Journal of Public Opinion Research*, vol. 17, no. 3, pp. 324–345, 2005.

[169] D. Keltner, K. Oatley, and J. M. Jenkins, *Understanding Emotions*, Wiley Hoboken, NJ, New Jersey, 2014.

[170] E. Diener and R. A. Emmons, "The independence of positive and negative affect," *Journal of Personality and Social Psychology*, vol. 47, no. 5, pp. 1105–1117, 1984.

[171] R. J. Larsen and E. Diener, *Promises and problems with the circumplex model of emotion*, Sage Publications, Inc, 1992.

[172] R. R. McCrae and P. T. Costa Jr., "Adding liebe und arbeit: the full five-factor model and well-being," *Personality and Social Psychology Bulletin*, vol. 17, no. 2, pp. 227–232, 1991.

[173] C. D. Broad, "Emotion and sentiment," *The Journal of Aesthetics and Art Criticism*, vol. 13, no. 2, pp. 203–214, 1954.

[174] A. Tellegen, *Personality traits: Issues of definition, evidence, and assessment*, University of Minnesota Press, 1991.

[175] J. J. Gross, S. K. Sutton, and T. Ketelaar, "Relations between affect and personality: support for the affect-level and affective-reactivity views," *Personality and Social Psychology Bulletin*, vol. 24, no. 3, pp. 279–288, 1998.

[176] D. Watson and L. A. Clark, "Negative affectivity: the disposition to experience aversive emotional states," *Psychological Bulletin*, vol. 96, no. 3, pp. 465–490, 1984.

[177] F. Rogers, "Personality and individuality," *The North American Review*, vol. 214, no. 791, pp. 514–517, 1921.

[178] L. Yue, W. Chen, X. Li, W. Zuo, and M. Yin, "A survey of sentiment analysis in social media," *Knowledge and Information Systems*, vol. 60, no. 2, pp. 617–663, 2019.

[179] Y. Wang, Z. Chen, and C. Fu, "Synergy masks of domain attribute model DaBERT: emotional tracking on time-varying virtual space communication," *Sensors*, vol. 22, no. 21, p. 8450, 2022.

[180] K. Vishnubhotla and S. M. Mohammad, "Tweet emotion dynamics: emotion word usage in tweets from US and Canada," 2022, https://arxiv.org/abs/2204.04862.

[181] S. Mohammad, "Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words," in *Proceedings of the 56th annual meeting of the association for computational linguistics (volume 1: Long papers)*, pp. 174–184, Melbourne, Australia, July 2018.

[182] R. Krishna and C. M. Prashanth, "Finding context-based influencers on Twitter," *SN Computer Science*, vol. 5, no. 1, 2024.

[183] R. O. Lutkenhaus, J. Jansz, and M. P. A. Bouman, "Tailoring in the digital era: stimulating dialogues on health topics in collaboration with social media influencers," *Digital Health*, vol. 5, article 205520761882152, 2019.

[184] S. B. Daily, M. T. James, D. Cherry et al., "Affective computing: historical foundations, current applications, and future trends," *Emotions and Affect in Human Factors and Human-Computer Interaction*, vol. 1, pp. 213–231, 2017.

[185] C. Zhang, A. Gupta, X. Qin, and Y. Zhou, "A computational approach for real-time detection of fake news," *Expert Systems with Applications*, vol. 221, article 119656, 2023.

[186] D. Matravers, *Art and Emotion*, Oxford University Press, Oxford, England, 2001.

[187] Y. Treadwell, "Humor and creativity," *Psychological Reports*, vol. 26, no. 1, pp. 55–58, 1970.

[188] J. Morreall, "Humor and emotion," *American Philosophical Quarterly*, vol. 20, no. 3, pp. 297–304, 1983.