

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp



Counterfactual thinking as a prebunking strategy to contrast misinformation on COVID-19^{\star}



Mauro Bertolotti^{*}, Patrizia Catellani

Department of Psychology, Catholic University of Milan, Italy

ARTICLEINFO	A B S T R A C T
Keywords: Misinformation Fake news Counterfactual thinking Prebunking Conspiracy mentality	Given the complexity of contrasting the spread of fake news and conspiracy theories, past research has started investigating some novel pre-emptive strategies, such as inoculation and prebunking. In the present research, we tested whether counterfactual thinking can be employed as a prebunking strategy to prompt critical consider- ation of fake news spread online. In two experiments, we asked participants to read or generate counterfactuals on the research and development of COVID-19 treatments, and then to evaluate the veridicality and plausibility of a fake news headline related to the topic. Participants' conspiracy mentality was also measured. Among participants with higher levels of conspiracy mentality, those exposed to counterfactual prebunking rated the fake news headline less plausible than those in the control condition (Study 1) and those exposed to another type of prebunking, that is, forewarning of the existence of misinformation (Study 2). The counterfactual prebunking strategy also induced less reactance than forewarning. Discussion focuses on the development of new strategies to prevent the spread of misinformation, and the conditions under which these strategies may be successful.

Recent research on the spread of misinformation has identified several psychosocial factors that make individuals more likely to believe in fake news (Bronstein, Pennycook, Bear, Rand, & Cannon, 2019; Pennycook & Rand, 2020; Tappin, Van Der Leer, & McKay, 2017; Van Prooijen, 2019). Furthermore, some studies have highlighted a concerning tendency to persist in one's belief in fake news even after new information and factual evidence of their falsehood is provided (Nyhan & Reifler, 2010; Walter & Tukachinsky, 2020), as certain individuals actively counter-argue corrections, or rationalize the inconsistencies between facts and their beliefs.

One of the processes involved in the persistence of belief in misinformation is counterfactual thinking (Roese, 1997), which allows individuals to uphold factually incorrect beliefs by mentally simulating the conditions under which those false facts *could have been true* (Effron, 2018). By using this strategy, believers of conspiracy theories and fake news can effectively insulate their beliefs from reality, uphold their opinions, and even justify subsequent behaviours, such as spreading them to other people.

In this paper, for the first time we propose that counterfactual thinking may also function as a pre-emptive strategy to make individuals *more* cautious of appealing but ultimately dubious news they may encounter. A counterfactual-based prebunking intervention may induce individuals to critically assess information regardless of its veridicality status, thus contemplating the content of fake news merely as a possible alternative to reality. We also expect this type of strategy to be perceived as less blunt and confrontational than other prebunking strategies, such as directly forewarning individuals of the existence of misinformation, thus reducing psychological reactance (Brehm & Brehm, 2013) and increasing its likelihood to succeed (Catellani & Bertolotti, 2014).

We ran two experimental studies, in which we asked participants to engage (or not) in counterfactual thinking before exposing them to a headline containing a piece of fake news regarding the COVID-19 pandemic. In both studies, we investigated participants' conspiracy mentality as a potential moderator of the effect of the counterfactual priming.

1. Prebunking strategies to correct misinformation and conspiracy theories

The most frequently used strategy to counter misinformation is to "debunk" it through fact-checking, rebuttals, or counterarguments (see Chan, Jones, Hall Jamieson, & Albarracín, 2017, for a meta-analysis).

https://doi.org/10.1016/j.jesp.2022.104404 Received 12 March 2022: Received in revised for

Received 12 March 2022; Received in revised form 8 August 2022; Accepted 30 August 2022 Available online 7 September 2022 0022-1031/© 2022 Elsevier Inc. All rights reserved.

 $^{^\}star\,$ This paper has been recommended for acceptance by Dr Nicholas Rule

^{*} Corresponding author at: Department of Psychology, Catholic University of Milan, Largo Gemelli, 1, I-20123 Milan, Italy. *E-mail address:* mauro.bertolotti@unicatt.it (M. Bertolotti).

Despite its seemingly straightforward application, the debunking approach has some drawbacks limiting its real-world effectiveness (Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012; Swire-Thompson, DeGutis, & Lazer, 2020; Wood & Porter, 2019). Recent research has thus explored the idea of dealing with misinformation pre-emptively, rather than correcting it afterwards (Compton, 2013; Jolley & Douglas, 2017). This strategy, aptly named "prebunking", attempts to anticipate misinformation by making potential targets aware of its existence, and providing them the necessary knowledge to recognize and counterargue it (Banas & Miller, 2013; Jolley & Douglas, 2017). Such approach is based on classic research on the so-called inoculation theory (McGuire, 1970; McGuire & Papageorgis, 1961), which proposed the idea of exposing individuals to a weakened and controlled version of a persuasive argument to prepare them for future exposure to full-blown persuasive attempts, in the same way as a vaccine shot trains the immune system to recognize and combat future pathogens.

Recent research has proposed an "active" approach to prebunking, where individuals are not only made aware of misinformation before being exposed to it, but actively engage in reading and evaluating misinformation content, familiarizing with its format, style, and persuasive potential through interactive online games (*Bad News*, Roozenbeek & Van der Linden, 2019; *Go Viral!*, Basol et al., 2021). The findings of these studies indicate that active prebunking improves participants' ability and confidence in identifying fake news.

Prebunking strategies can be effective in correcting misinformation, but some relevant limitations remain. Although they provide the necessary awareness and skill to detect misinformation, we have so far little evidence (Lewandowsky & Van Der Linden, 2021) that the cognitive resources acquired through prebunking are effectively activated when individuals are exposed to fake news in their actual informational environment (Lee & Chyi, 2014). To overcome this limitation, it may be useful to leverage on the same processes that individuals spontaneously engage in relation with misinformation. In our research, we focused on counterfactual thinking as a potential strategy of this type.

2. Counterfactual thinking: A support to falsehoods or a prebunking strategy?

Counterfactual thinking (Roese, 1997) is a form of mental simulation in which a fact or event is undone or mutated by hypothetically changing its antecedents, to obtain a different, usually more desirable, outcome. As far as we know, counterfactual thinking has never been investigated as a debunking or prebunking strategy, but, on the contrary, it has been studied as a process involved in individuals' motivated reasoning aimed at supporting belief in fake news and conspiracist narratives (Effron, 2018). In those studies, participants first read actual facts regarding political events and candidates, and then evaluated false statements that contradicted those facts. Results showed that participants were more likely to consider such falsehoods plausible, and condone their diffusion, when they were also presented counterfactual messages stating that the falsehood could have been true. More precisely, participants who were exposed to counterfactual messages considered falsehood aligned with their political preferences more plausible, and thus closer to reality, than participants in the respective control conditions. Past research has also shown that engaging in mental simulation can provide individuals with a viable justification to defend unsupported claims (Shalvi, Dana, Handgraaf, & De Dreu, 2011). When confronted with the fact that something they believe is not true, people may resort to considering how it could have been true, or it could still become true in the future (Helgason & Effron, 2022) to maintain some degree of belief and avoid undergoing the cognitively and motivationally costly process of changing their mind.

Assuming the Reflection and Evaluation Model of Comparative Thinking (Markman & McMullen, 2003, 2005) as a theoretical reference, one may infer that the counterfactual prompts employed in the studies described above activated a *counterfactual reflection mode* in

participants, that is, the a posteriori simulation of alternative outcomes to a known reality. This effectively allowed them to hang on to a desirable untruth (Petrocelli, Seta, & Seta, 2013). We hypothesized, conversely, that a prebunking strategy inducing counterfactual thinking a priori (i.e., before exposure to fake news) would prompt a counterfactual evaluation mode, that is a comparison between reality and simulated alternatives which induces individuals to engage in a flexible but careful consideration of reality and its potential alternatives (Effron, 2018, p. 742). In our specific case, counterfactual thinking would prompt individuals to funnel misinformation into a "What if..." mental domain, and consider the possibility and plausibility of different scenarios before coming to a conclusion on its veridicality (Byrne, 2002; De Brigard, Henne, & Stanley, 2021). This hypothesis is supported by the known association of counterfactuals with self-regulation and problem solving (Epstude & Roese, 2008; Roese & Epstude, 2017), as counterfactual thinking allows individuals to identify problematic aspects of a past situation and prepare adaptive responses to them in the future. Thus, counterfactual thinking promotes analytical thinking, which has been found to be negatively associated with the perceived accuracy of fake news in several studies (Bago, Rand, & Pennycook, 2020; Bronstein, et al., 2019; Swami, Voracek, Stieger, Tran, & Furnham, 2014; see Sindermann, Cooper, & Montag, 2020, for a review). The analytical and critical information processing style elicited by counterfactual thinking might therefore help individuals in assessing the claims contained in fake news and conspiracy theories, evaluating their internal consistency before committing to believing or rejecting them.

In sum, compared to other prebunking strategies, which tend to focus the audience on their cognitive shortcomings, counterfactual thinking may have the advantage of mobilising individuals' own cognitive abilities, activating a counterfactual evaluation mode, and promoting analytical thinking. This effect, however, might not be the same for all individuals. In the context of research on belief in fake news it is particularly important to understand how counterfactual thinking might affect individuals with a conspiracy mentality.

3. Targeting individuals with a conspiracy mentality

Past research has focused on conspiracy mentality (Bruder, Haffke, Neave, Nouripanah, & Imhoff, 2013) as a relevant individual dimension in the tendency to believe in misinformation. Conspiracy mentality is "the general propensity to subscribe to theories blaming a conspiracy of ill-intending individuals or groups for important societal phenomena" (Bruder et al., 2013, p. 2). Some studies have shown that individuals with a high conspiracy mentality are considerably more likely than average to believe in fake news disseminated online (Mancosu, Ladini, & Vassallo, 2021; Swami et al., 2017), and more likely to distrust institutional and mainstream media sources, thus making them less susceptible to communication telling them what to believe or not (as prebunking and debunking messages tend to do).

Counterfactual thinking has been associated with conspiracist ideation (Galinsky, Liljenquist, Kray, & Roese, 2005; Moscovici, 2020) and some authors specifically identified conspiracist thinking as a form of biased counterfactual thinking (THUNCing, or Thinking in Unreflexive Counterfactuals, Lewandowsky, Lloyd, & Brophy, 2018; Lewandowsky et al., 2015). Therefore, individuals with a conspiracist mentality are usually motivated not to take information at face value, explore different accounts of events, and contemplate alternative explanations (other than the "official truth", Uscinski, 2018). As such, a prebunking strategy precisely prompting individuals to think about multiple potential alternatives might be particularly suitable for those with a conspiracist mentality.

Furthermore, as individuals with a conspiracy mentality tend to distrust mainstream media sources (Anthony & Moulding, 2019; Bode & Vraga, 2018; Imhoff & Bruder, 2014), they are also very likely to react negatively to attempts to correct their beliefs and explanations of events by such sources. This makes the typical debunking and prebunking

messages more likely to produce reactance among these people, rather than to change their mind (Ecker et al., 2022; Hornsey & Imani, 2004; Rabinovich & Morton, 2010). The hypothetical format of counterfactual thinking, however, makes it less blunt and injunctive than factual or direct messages (Catellani, Bertolotti, & Covelli, 2013; Fiedler & Mata, 2013), and thus less likely to trigger reactance, as past research has shown in the context of contentious political exchanges (Bertolotti & Catellani, 2018; Catellani & Bertolotti, 2014), and judicial decisionmaking (Catellani, Bertolotti, Vagni, & Pajardi, 2021; Tal-Or, Boninger, Poran, & Gleicher, 2004; Wong, 2010).

In sum, we expect individuals with a conspiracist mentality to be relatively more receptive to counterfactual-based interventions than other individuals, and less likely to "close out" from persuasive attempts presented in a counterfactual format than from those presented as direct messages.

4. The present research

In two studies, we investigated counterfactual thinking as a potential prebunking strategy by testing the effects of counterfactual thoughts on participants' belief in fake news presented thereafter. The first study aimed at exploring the effects of two counterfactual interventions, whereas the second study aimed at confirming our initial findings and extending them through the comparison with an alternative prebunking approach.

In Study 1, we asked participants to either read a counterfactual message or generate counterfactual thoughts regarding the development of pharmaceutical treatments for COVID-19, which has been the topic of intense misinformation since the beginning of the pandemic and during the vaccine rollouts in 2021. We then showed participants some headlines and asked to evaluate them in terms of veridicality (i.e., whether they were real or fake news) and plausibility. In this initial study, we compared a prebunking intervention based on reading counterfactual statements with one based on reading and generating counterfactual thoughts. These two approaches differ in some relevant regards. First, in the amount of cognitive effort required to complete each task, as the generation of counterfactual thoughts entails additional time and involvement compared to simply reading a series of statements. Second, the counterfactuals used in the messages were essentially semantic counterfactuals (Revlin, Cate, & Rouss, 2001; Roese & Epstude, 2017), based on common knowledge on the issue of COVID-19 and related treatments, whereas self-generated counterfactual thoughts may have included also episodic counterfactuals (De Brigard, Addis, Ford, Schacter, & Giovanello, 2013), based on participants' individual experience with the issue. As we could not control the type and content of the counterfactuals generated by participants, we expected more heterogeneous, and therefore weaker, effects in the counterfactual generation condition.

In Study 2, we compared the exposure to a counterfactual message with another type of prebunking message, simply forewarning participants of the existence of fake news on the topic under discussion. We also measured participants' reaction to these two types of pre-emptive strategies.

In both studies, we considered participants' conspiracy mentality (Bruder et al., 2013) as a potential moderator of the effects of counterfactual prebunking strategies, given the greater tendency of individuals with a high conspiracy mentality to believe in fake news (Anthony & Moulding, 2019; Bode & Vraga, 2018), but also their disposition towards considering multiple alternative explanations for events (Lewandowsky et al., 2015), and their susceptibility to psychological reactance when approached with other types of correction (Ecker et al., 2022).

In Study 1, which was mainly exploratory, we formulated the following research questions:

Research Question 1. Are participants with high conspiracy mentality exposed to a counterfactual message less likely to consider the headline plausible (RQ1a) and more likely to recognize it as fake (RQ1b), compared to when they are in a control condition?

Research Question 2. Are participants with high conspiracy mentality asked to generate counterfactual thoughts less likely to consider the headline plausible (RQ2a) and more likely to recognize it as fake (RQ2b), compared to those in a control condition?

Research Question 3. Are the above differences reduced (or even annulled) for participants with low conspiracy mentality?

We expected participants with high conspiracy mentality to particularly benefit from counterfactual prebunking, as they are both more susceptible to the dubious claims contained in the headline (Pennycook & Rand, 2020; Ross, Rand, & Pennycook, 2021) and more inclined to process information in this peculiar way (Galinsky et al., 2005; Lewandowsky et al., 2015). This effect, however, might be attenuated in the counterfactual generation condition, as participants may generate counterfactuals that are not aligned with the intended prebunking function (e.g., personally relevant episodic counterfactuals, or counterfactuals based on conspiracy theories).

In Study 2, we sook confirmation of our findings comparing the effects of counterfactual prebunking not only with a control condition but also with another form of prebunking message, that is, forewarning. We formulated the following hypotheses:

Hypothesis 1. Participants with high conspiracy mentality exposed to a counterfactual message are less likely to consider the fake news headline plausible than the same type of participants exposed to a forewarning prebunking message, or in the control condition (H1a), whereas such differences are reduced (or even annulled) for participants with low conspiracy mentality (H1b).

Hypothesis 2. Participants with high conspiracy mentality exposed to a counterfactual message are more likely to recognize the headline as fake than the same type of participants exposed to a forewarning prebunking message, or in the control condition (H2a), whereas such differences are reduced or even annulled for participants with low conspiracy mentality (H2b).

In Study 2, we also aimed at assessing whether counterfactual prebunking would trigger less reactance than forewarning. We expected that this would be the case because a direct warning about the spread of fake news (as in the case of the forewarning message) is likely to trigger reactance, particularly among those who are inclined to believe in conspiracy theories. These individuals are motivated to endorse ideas that deviate from majority opinion (Imhoff & Erb, 2009; Imhoff & Lamberty, 2017), and might therefore react negatively to messages explicitly denouncing such ideas. A counterfactual message speculating on alternative realities, conversely, might appeal to these individuals, as it does not explicitly counter those beliefs (Fiedler & Mata, 2013), and it mimics the common rhetoric style used to propagate conspiracy theories (the "just asking questions" argument, Starbird et al., 2016). We therefore formulated the following hypothesis:

Hypothesis 3. Participants with a high conspiracy mentality exposed to a counterfactual message are less likely to express reactance to the message than those exposed to a forewarning prebunking message (H3a). Such difference is reduced or annulled among participants with low conspiracy mentality (H3b).

If our expectations were corroborated, our results would show that counterfactual thinking may be employed as a novel prebunking strategy to contrast misinformation, particularly among individuals with a tendency to believe in conspiracy theories.

5. Study 1

In Study 1, we investigated two types of counterfactual prebunking strategies, namely the exposure to a counterfactual message related to the topic of the misinformation, and the generation of counterfactual thoughts by participants themselves. We compared participants' veridicality and plausibility judgements regarding the fake news headline with those made by participants in a control condition, neither reading nor generating counterfactual thoughts. Below, we report all data exclusions, manipulations, and measures in the study.

5.1. Method

5.1.1. Participants

Participants were recruited online, after being contacted through direct contact or social media by Italian students. A total of 1117 people were contacted, and 952 participants (52.9% females, 45.6% males, age M = 39.9, SD = 16.4) completed the full questionnaire and were randomly assigned to either the counterfactual message condition (n =331), the counterfactual generation condition (n = 294), or the control condition (n = 327). In the counterfactual generation condition, a small number of participants failed to generate a valid counterfactual thought. They either left the space blank, entered non-answers (e.g., "I don't know"), or entered non-counterfactual thoughts (e.g., "I think the best possible result was achieved", or "I don't think better results could have been obtained"). These participants were excluded from subsequent analyses, reducing the final number of participants in the counterfactual generation condition to n = 263. Excluded participants did not differ from the others in terms of age, t(290) = 0.37, p = .713, ¹ gender, $\chi^2(2, N)$ $(294) = 0.621, p = .733, or education level, <math>\chi^2(6, N = 294) = 1.338, p$ = .969. A sensitivity power analysis using GPower 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007) was conducted for the interaction effect between the experimental manipulation and participants' conspiracy mentality, in a multiple regression with 5 total predictors (two dummy variables representing the experimental conditions, a continuous moderator, and two interaction terms). The analysis indicated that with a significance level of p = .05 (two-tailed), and a power of $\beta = 0.80$, the final sample size provided sufficient power to detect an effect with $f^2 =$ 0.008, corresponding to an R-squared value change of $\Delta R^2 = 0.0079$.

5.1.2. Procedure

In the two experimental conditions, participants were first invited to read a brief text on the current state of research on pharmaceutical treatments for COVID-19:

"In recent months research on drugs to treat the symptoms of COVID-19 has led to the development of several treatment protocols. Their current effectiveness and reliability, however, make it premature to speak of a real cure for this disease."

In the counterfactual message condition, the text was followed by a second part containing three separate upward counterfactual thoughts. The message was the following:

"Many think that research in this area would have obtained better results, if only some things had gone differently. For example, some think that if pharmaceutical companies had allocated more money to research on drug therapies, today we would have more medicines available to treat the symptoms of COVID-19. Others think that if states had not directed all their efforts only to the development of vaccines, we would now also have tools to treat as well as prevent COVID-19."

In the counterfactual generation condition, the counterfactual message described above was followed by an additional prompt inviting participant to generate similar thoughts on the issue:

"Now try to think yourself what could have gone differently by completing the following sentence. Research on drugs to treat COVID-19 would have had better results if...".

Participants were then presented with four illustrated headlines presenting news on the issue of COVID-19. They were told that each

headline had a 50% chance of being true and a 50% chance of being fake. Whereas the first three headlines were only generically related to the issue and served as filler stimuli, the final one specifically regarded the topic discussed in the manipulation task, that is pharmaceutical treatments for COVID-19, as it read as follows: "From research on plants, a new molecule to treat COVID-19 symptoms" (Fig. 1). Participants then proceeded to a following page of the questionnaire, where they were told that the last headline reported a fake news and were asked some additional follow-up questions.

5.1.3. Measures

5.1.3.1. Headline prior knowledge. After reading each headline, participants were asked whether they had ever heard or seen that piece of news. The response options were "Yes", "No", or "Don't know".

5.1.3.2. Headline veridicality. Participants were also asked whether they thought the headline was true or fake. The response options were again "Yes", "No", or "Don't know".

5.1.3.3. *Plausibility*. After reading the target headline, participants were asked to indicate to what extent they thought it was reliable, verisimilar, and plausible, using a 7-point scale ranging from "Not at all" (1) to "Very" (7). The three item scores were averaged into a single plausibility index (Cronbach's $\alpha = 0.920$).

5.1.3.4. Conspiracy mentality. In the following section of the questionnaire, participants were asked their agreement with 5 statements, adapted from the conspiracy mentality scale (Bruder et al., 2013), using a scale from 1 ("Completely disagree") to 7 ("Completely agree"). The item scores were later averaged into a single conspiracy mentality index ($\alpha = 0.815$). As a preliminary check, we tested whether participants in the three experimental conditions differed in conspiracy mentality, finding no significant differences among the three groups, F(2, 949) =0.93, p = .395. Means, standard deviations, and zero-order correlations among the main variables are reported in Table 1 (upper pane).

5.1.3.5. Other variables. Participants' basic sociodemographic characteristics were recorded, and are reported in the Supplementary Materials. A few additional questions were included in the questionnaire for exploratory purposes. Measures of the ethicality of spreading fake news (Effron, 2018) and behavioural intentions towards online misinformation content (Effron & Raj, 2020) are analysed in the Supplementary Materials. Other measures of individual and collective risk perception from the pandemic, political orientation, and populism were not used in the present study.

5.2. Results and discussion

5.2.1. Headline prior knowledge

Most participants reported not having read the target fake headline before (87.1%), in similar proportions across the three experimental conditions, $\chi^2(4, N = 915) = 4.52$, p = .340. No significant differences were found for the other filler headlines either, $\chi^2(4, N = 915) < 5.57$, p > .234.

5.2.2. Headline plausibility

We then analysed participants' plausibility judgements across the experimental conditions and different levels of conspiracy mentality. We ran a regression model using PROCESS (Hayes, 2018, Model 1), with the plausibility index score as the dependent variable and two dummy variables, representing the counterfactual message condition and the counterfactual generation condition as the main predictors. The conspiracy mentality score was entered in the model as an additional predictor, as well as the two interaction terms with the above-mentioned

¹ N = 5 participants did not report their age (n = 1 in the control condition, n = 2 in the counterfactual message condition, and n = 2 in the counterfactual generation condition).

Pediatric study finds: wearing face masks for hours reduces children's cognitive skills



Supplements to cure COVID-19: lactoferrin, quercetin, and zinc reduce symptom strength and duration



COVID-19 treatment trial: hyperimmune plasma prevents death and reduces symptoms



From research on plants, a molecule to cure the symptoms of COVID-19



Fig. 1. Fake News Headline Stimuli Used in Studies 1 & 2 (English Translation of the Original Italian Text; Target Stimulus Highlighted).

 Table 1

 Descriptive statistics for the main variables (Studies 1 & 2).

Study 1						
	N M			SD	1	2
1. Headline Plausibility	915 3.08		3	1.45	-	0.055
2. Conspiracy Mentality	915	4.06		1.34		-
Study 2						
	Ν	М	SD	1	2	3
1. Headline Plausibility	494	3.58	1.23	_	0.121*	0.216**
2. Conspiracy Mentality	494	3.59	1.26		-	0.319**
3. Psychological Reactance ^a	330	2.73	1.23			_

Notes: * *p* < .05; ** *p* < .001.

^a not measured in the control condition (N = 164).

dummy variables. The analysis showed no main effect of the counterfactual message condition, B = -0.14, SE = 0.11; t(915) = 1.27, p =.204, 95% CI [-0.37; 0.08], nor of the counterfactual generation condition, B = -0.02, SE = 0.12; t(915) = 0.19, p = .853, 95% CI [-0.26; 0.21], whereas a direct, positive effect of conspiracy mentality was found, B = 0.16, SE = 0.06; t(915) = 2.59, p = .010, 95% CI [0.04; 0.28], indicating that participants with higher conspiracy mentality regarded the headline as more plausible than participants with lower conspiracy mentality. We then found a significant interaction effect between the counterfactual message condition and conspiracy mentality, B = -0.19, SE = 0.09; t(915) = 2.17, p = .031, 95% CI [-0.36; -0.02]. Inspecting conditional effects of the experimental conditions at high (+1 SD) and low (-1 SD) levels of conspiracy mentality (Fig. 2), we found that among the former, exposure to a counterfactual message significantly reduced the headline plausibility as compared to the other two conditions, B_{high} = -0.39, SE = 0.16; t(915) = 2.39, p = .017, 95% CI [-0.72; -0.07], whereas no significant difference was found among participants with lower levels of conspiracy mentality, $B_{low} = 0.11$, SE = 0.16; t(915) =

0.67, p = .503, 95% CI [-0.21; 0.42]. A smaller and non-significant interaction effect was found also for the counterfactual generation condition, B = -0.11, SE = 0.09; t(915) = 1.27, p = .204, 95% CI [-0.29; 0.06], omnibus moderation effect $\Delta R^2 = 0.005$, p = .093.

5.2.3. Headline veridicality

Participants in the three experimental conditions did not differ in their evaluation of the target headline veridicality, with similar shares of participants recognizing it as fake (44.3% in the control condition, 41.1% in the counterfactual message condition, and 41.1% in the counterfactual generation condition), a small proportion of them indicating it was true (14.7%, 11.5%, and 12.2%, respectively), and the rest indicating that they did not know (41.0%, 47.4%, and 46.8%, respectively), $\chi^2(4, N = 915) = 3.76$, p = .440. We ran a logistic regression model using PROCESS (Hayes, 2018, Model 1) to test whether the likelihood of recognizing the headline as fake was influenced by the interaction between the experimental condition and participants' conspiracy mentality. The dependent variable was recoded into a dichotomic variable, with the values of 1 attributed to participants who recognized the headline as fake, and 0 to those who reported it being true or did not know, respectively. The analysis showed no main effect of the counterfactual message, B = -0.17, SE = 0.16; Z(915) = 1.06, p =.290, 95% CI [-0.48; 0.14], nor of the counterfactual generation conditions, *B* = 0.17, *SE* = 0.17; *Z*(915) = 0.99, *p* = .323, 95% CI [-0.50; 0.16], whereas a negative effect of conspiracy mentality only approached significance, B = -0.16, SE = 0.09; Z(915) = 1.88, p = .060, 95% CI [-0.33; 0.01], indicating that participants with higher conspiracy mentality were less likely to spot the fake headline than participants with lower conspiracy mentality. We then found a significant interaction effect between the counterfactual message condition and conspiracy mentality, *B* = 0.25, *SE* = 0.12; *Z*(915) = 2.08, *p* = .037, 95% CI [0.01; 0.49], mirroring our previous finding on plausibility, whereas no correspondent interaction effect emerged between the counterfactual generation condition and conspiracy mentality, B = 0.17, SE = 0.12; Z (915) = 1.34, p = .180, 95% CI [-0.08; 0.41].



Fig. 2. Headline Plausibility as a Function of Experimental Condition and Participants' Conspiracy Mentality (Study 1).

5.3. Discussion

Taken together, these findings provided evidence to answer our RQ1, as they showed that among participants with high conspiracy mentality the counterfactual message was successful in reducing the plausibility and veridicality attributed to the fake headline compared to the control condition. This was only partially true for counterfactual generation (RQ2), as the observed trend appeared to be weaker and nonsignificant. This might be the case because some participants evoked hypothetical alternatives to reality that were not functional to a critical scrutiny of the headlines they later read, as compared to the semantic counterfactuals provided in the counterfactual message condition. Finally, no such effects emerged among participants with lower conspiracy mentality (RQ3).

6. Study 2

In Study 2, we followed up on the findings of Study 1 and compared counterfactual prebunking messages with a more standard prebunking approach (i.e., forewarning). We expected reduced headline plausibility and veridicality in the counterfactual prebunking condition than in the simple prebunking and control conditions, particularly in the case of participants with a high level of conspiracy mentality. Furthermore, we expected counterfactual communication to trigger overall less reactance than the forewarning approach. We report all data exclusions, manipulations, and measures in the study.

6.1. Method

6.1.1. Participants

Participants were contacted through the Prolific online platform, where they were asked to participate in a study on "online sources of information". Of the initial 498 participants, a total of 494 Italian-speaking participants (49.6% males, 49.0 females, age M = 28.5, SD = 9.6) completed the full questionnaire (with n = 2 participants dropping out after beginning in the control condition and n = 2 in the counterfactual message condition). No participants were excluded due to failure in completing the task. A sensitivity power analysis was conducted for the hypothesized (H1) interaction effect between the experimental manipulation and participants' conspiracy mentality, with the

same criteria used in Study 1. The analysis indicated that our sample size provided sufficient power to detect an effect with $f^2 = 0.016$, corresponding to an $R^2 = 0.0157$. No participants were excluded from the analyses in this study.

6.1.2. Procedure

The procedure was similar to the one employed in Study 1, with an initial brief text describing the current state of research on drug treatments for COVID-19. Then, participants in the counterfactual message condition read the same counterfactual statements presented in Study 1. This time, however, the message ended with an additional statement remarking that: "Naturally, these are just hypotheses on how things could have been". Participants in the prebunking message condition read a message aimed at forewarning participants regarding the existence of fake news on the issue of COVID-19 treatments. The message was the following:

"A lot of fake news have been spread online on this topic. For example, some argue that we do not yet have medicines to treat the symptoms of COVID-19 because research on drug therapies has been neglected for economic interests. Others think that research has proceeded slowly because trials and authorization procedures have been hindered. These reports are based on inaccurate, incomplete or often completely invented elements."

In the control condition, participants directly proceeded to the next page after reading the initial text, without reading any additional message.

Participants were then presented with the same headlines used in Study 1 and were asked some follow-up questions on the last headline containing a false claim on a drug treatment for COVID-19.

6.1.3. Measures

6.1.3.1. Headline prior knowledge and veridicality. As in Study 1, participants reported whether they had already seen each headline and whether they thought it was true or fake.

6.1.3.2. *Plausibility*. After reading the last headline, participants were further asked to indicate to what extent they thought it was reliable, verisimilar, and plausible (Cronbach's $\alpha = 0.881$), using the same 7-point scale ranging used in Study 1.

6.1.3.3. Conspiracy mentality. In the final section of the questionnaire, participants answered the same conspiracy mentality scale used in Study 1 ($\alpha = 0.828$). Participants did not differ in conspiracy mentality across the three experimental conditions, F(2, 491) = 1.97, p = .141.

6.1.3.4. *Reactance.* In the counterfactual message and prebunking message conditions, participants first evaluated the initial message, with the following four items adapted from Shen and Dillard (2005): "The text I have just read... tried to constrain my freedom of thought, aimed at influencing my opinion, tried to manipulate me, tried to pressure me". Responses were recorded on a 7-point scale ranging from "Not at all" (1) to "Very" (7), and later averaged into a single reactance index ($\alpha = 0.846$). Means, standard deviations, and zero-order correlations among the main variables are reported in Table 1 (lower pane).

6.1.3.5. Other variables. As in Study 1, participants' gender, age, education level, and profession were recorded. The ethicality of spreading fake news and behavioural intentions towards them were measured as well. Explorative analyses on these variables are reported in the Supplementary Materials.

6.2. Results and discussion

6.2.1. Headline plausibility

We first analysed participants' plausibility judgements in a multiple regression model with two dummy variables representing the counterfactual message condition and the prebunking message condition as the main predictors. The conspiracy mentality and the two interaction terms with the counterfactual and prebunking message conditions were entered in the model to test for the expected moderation effect. The analysis showed no main effects of the two experimental conditions, namely the counterfactual message, B = -0.10, SE = 0.14; t(488) = 1.46, p = .146, 95% CI [-0.46; 0.07], and the prebunking message condition, B = -0.08, SE = 0.14; t(488) = 0.56, p = .577, 95% CI [-0.34; 0.19], whereas a main effect of conspiracy mentality was found, B = 0.20, SE = 0.07; t(488) = 2.82, p = .005, 95% CI [0.06; 0.35], replicating the positive association between conspiracy mentality and plausibility attribution to the headline found in Study 1. We then found a significant interaction effect between the counterfactual message

condition and conspiracy mentality, B = -0.21, SE = 0.11; t(488) =2.05, p = .041, 95% CI [-0.43; -0.01], again replicating what we found in Study 1. In particular, simple comparisons revealed that among participants with high conspiracy mentality exposure to a counterfactual message significantly reduced the headline plausibility as compared to the other two conditions, $B_{high} = -0.47$, SE = 0.19; t(488) = 2.54, p =.011, 95% CI [-0.84; -0.11], whereas this was not the case among participants with low conspiracy mentality, $B_{low} = 0.08$, SE = 0.20; t (488) = 0.40, p = .689, 95% CI [-0.31; 0.46] (Fig. 3). The interaction effect with the prebunking message condition was not significant, B =-0.06, SE = 0.11; t(488) = 0.59, p = .556, 95% CI [-0.27; 0.14], omnibus moderation effect $\Delta R^2 = 0.009$, p = .115. This results therefore corroborated our H1a regarding the differential effects of the two types of messages on participants with higher conspiracy mentality, and our H1b regarding the relative similarity of the effects in the case of participants with lower conspiracy mentality.

6.2.2. Headline veridicality

As in Study 1, participants in the three experimental conditions did not differ in their evaluation of the target headline veridicality, $\chi^2(4, N = 494) = 2.65$, p = .618. The same logistic regression model tested in Study 1 was ran in this case. The analysis showed no main effects of the two experimental conditions, namely the counterfactual message, B =-0.01, SE = 0.26; Z(488) = 0.04, p = .967, 95% CI [-0.50; 0.52], and the prebunking message condition, B = -0.20, SE = 0.27; Z(488) =0.77, p = .440, 95% CI [-0.74; 0.32]. Furthermore, neither conspiracy mentality, nor its interactions with the experimental conditions, had significant effect on participants' recognition of the fake headline, Bs <0.11, ps > 0.633, thus not supporting our H2.

6.2.3. Reactance

This analysis was limited to participants in the two experimental conditions, excluding those in the control condition who did not answer the items regarding the prebunking message. The goal of this analysis was to test whether participants' reaction to the counterfactual prebunking message would differ from their reaction to the simple prebunking message.

In a regression model, with reactance as the dependent variable, we found a significant and strong effect of the experimental condition, B = 0.80, SE = 0.12; t(326) = 6.99, p < .001, 95% CI [0.58; 1.03], indicating



Fig. 3. Headline Plausibility as a Function of Experimental Condition and Participants' Conspiracy Mentality (Study 2).

that participants in the counterfactual message condition showed significantly less reactance (M = 2.26, SD = 1.34) than participants in the simple prebunking message condition (M = 3.01, SD = 0.86). No significant effect of conspiracy mentality emerged, B = -0.01, SE = 0.15; t(326) = 0.09, p = .926, 95% CI [-0.31; 0.28], but the predicted interaction effect was found, B = 0.24, SE = 0.09; t(326) = 2.49, p = .013, 95% CI [0.05; 0.42], showing that the difference between the two conditions was greater at higher levels (+1 SD above the mean) of conspiracy mentality, $B_{high} = 1.08$, SE = 0.16; t(326) = 6.71, p < .001, 95% CI [0.77; 1.41], as per our H3a, than at lower levels of conspiracy mentality (-1 SD), $B_{low} = 0.52$, SE = 0.16; t(326) = 3.17, p = .002, 95% CI [0.20; 0.84], as per our H3b. Overall, H3 was therefore corroborated.

6.3. Discussion

In sum, findings from Study 2 confirmed that the counterfactual message was successful in reducing the plausibility (but not the veridicality) attributed to the headline among participants with higher levels of conspiracy mentality, and further showed that this was not the case with the simple prebunking message, thus indicating a relative advantage of our approach compared to the more straightforward forewarning used in other studies in the past. Furthermore, we found that the counterfactual message induced comparatively less reactance than the simple prebunking message (again, particularly among participants with a high level of conspiracy mentality), confirming the suitability of this strategy to deal with sensitive topics, and suspicious, ill-disposed audiences.

7. General discussion

In the present research, we tested for the first time whether exposure to counterfactual messages regarding COVID-19 treatments can be used as a prebunking strategy to contrast fake news on the same issue. Results showed that individuals with a high conspiracy mentality are less likely to find plausible a fake headline when they are pre-emptively exposed to counterfactual messages proposing claims on the same issue as "What if..." scenarios. These participants also showed less reactance towards prebunking messages than when they were exposed to more direct forewarning.

Our findings are consistent with past research indicating that counterfactual thinking can promote critical and analytical thinking (Markman, Lindberg, Kray, & Galinsky, 2007) and problem solving (Roese & Epstude, 2017). The mere exposure to a counterfactual message primed individuals with relatively high levels of conspiracy mentality to make more cautious evaluations of the headline stimulus' plausibility (and, to some extent, veridicality). The selective effect on individuals with a high conspiracy mentality (i.e. with a tendency to distrust authorities, raise doubts about official and mainstream narratives, and attribute events to obscure and hidden agents; Bruder et al., 2013) might depend on their disposition to entertain counterfactual thoughts that question the reality of events and seek new potential explanations for them (Lewandowsky et al., 2015; Moscovici, 2020). So, in a certain sense, a counterfactualbased intervention to counter misinformation can be seen as an attempt to meet these people halfway down the "rabbit hole" (Uscinski, 2018), and engage in a type of speculation they are familiar with and willing to accept.

Our findings also indicate that counterfactual prebunking may provide an indirect approach to countering misinformation that may be useful with individuals impervious to other more direct strategies. Several studies have now established that conspiracy theories and misinformation serve multiple purposes for individuals (Douglas et al., 2019; Douglas, Sutton, & Cichocka, 2017), particularly when they deal with issues with strong personal relevance and central to one's social and political identity. Existential and social motives, in addition to epistemic ones, may cause considerable resistance when someone attempts to deny or correct misinformation (Ecker et al., 2022). Instead of directly targeting the veridicality, accuracy, and reliability of misinformation content, a counterfactual intervention frames such content as an hypothesis one can freely take in consideration before coming to a definitive conclusion. Such approach therefore provides a sort of middle ground that prevents individuals, and particularly those with a tendency to engage in conspiracist ideation, from closing out from attempts to change their mind. This finding is in line with past research on the persuasive properties of counterfactual communication, which have been previously investigated in other contexts (Bertolotti & Catellani, 2018; Catellani et al., 2021; Tal-Or et al., 2004; Wong, 2010). This makes our proposed strategy a promising tool to treat the problem exactly in the cases where it seems most difficult to deal with, as conspiracist individuals are also most likely to react negatively to other forms of interventions such as debunking or fact-checking.

Our research has some relevant limitations that future research might address, by further developing the experimental paradigm we employed, and by extending it to other domains of online communication. First of all, we found small effects (needing further replications with larger samples), which suggests that these specific counterfactual prebunking interventions may have only limited effectiveness. Our manipulation consisted in a short text introducing the issue of COVID-19 treatments, and either anticipating misinformation in a hypothetical format (i.e., "Many think that if..., then...") or prompting participants to generate similar thoughts themselves. By doing so, participants inevitably acquired some additional (albeit minimal) information compared to participants in the control condition, who were exposed directly to the headlines. Furthermore, the task was quite different from what social media users and online news consumers encounter in their daily browsing. Future research could therefore refine these manipulations, embedding them in a more realistic content (e.g., social media posts, or comments by other users) to improve their ecologic validity. On a similar note, participants in our studies were told that the headlines they were evaluating had a 50% chance of being fake news. This was done to make them doubt about the stimuli they were exposed to (so to avoid them taking the headline for true or fake with 100% certainty) and for ethical reasons (as presenting fabricated headlines as legitimate would have required an extensive and in-depth debriefing that was not possible in an online study). Such set up, which was present also in the control conditions of the two studies, might have partially obscured the comparative effectiveness of the prebunking interventions tested in Studies 1 & 2, as all participants had a reason to suspect that the headlines may have not been true. As for the content of the stimuli, we investigated the case of misinformation on COVID-19 treatments, as it was an issue that most participants would be aware of and probably quite interested in. The fake news headline furthermore made a hopeful, positive claim (i.e., that research on plants resulted in a cure for the disease), which we expected to attract participants' attention and motivate them to believe in it, as positive affect is often associated with heuristic and intuitive thinking (Bertolotti & Catellani, 2021; Forgas, 1995; Greifeneder, Bless, & Pham, 2011). Future research should explore the cognitive mechanisms underlying the effects we observed, measuring the extent to which participants effectively engage in counterfactual thinking after reading a counterfactual message, and testing different boundary conditions, e.g., manipulating the topic of misinformation, the subjective relevance and valence of its claims, the emotions they evoke (Martel, Pennycook, & Rand, 2020), and the role of conflict and partisanship (Osmundsen, Bor, Vahlstrup, Bechmann, & Bang Petersen, 2021; Vosoughi, Roy, & Aral, 2018).

To conclude, we explored for the first time the potential of counterfactual communication and counterfactual thinking as a strategy to pre-emptively contrast misinformation. Results indicate that this approach might be useful particularly to target individuals who are more prone to believing in such type of content, such as those with a conspiracy mentality. If applied to actual real-world scenarios, the approach we propose could contribute to clearing our informational environment from the fake news and misinformation that have plagued

M. Bertolotti and P. Catellani

it in recent years, not by chasing false claims and censoring them, but by providing users with a practical and mindful way of navigating the news.

Open practices

Data used in Studies 1 and 2 are available for download at htt ps://osf.io/47cmw/?view_only=1e3f0b862c5b4b02a4ae15635f3 ba9d3.

Funding

This research was funded by the Catholic University of Milan (grant line D.3.2 2020).

Declaration of Competing Interest

The authors hereby declare no conflict of interests.

Acknowledgements

We would like to thank Kai Sassenberg and two anonymous Reviewers for the very useful and insightful comments to the original version of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jesp.2022.104404.

References

- Anthony, A., & Moulding, R. (2019). Breaking the news: Belief in fake news and conspiracist beliefs. Australian Journal of Psychology, 71(2), 154–162. https://doi. org/10.1111/ajpy.12233
- Bago, B., Rand, D. G., & Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of Experimental Psychology: General*, 149(8), 1608–1613. https://psycnet.apa.org/doi/10.1037/x ge0000729.
- Banas, J. A., & Miller, G. (2013). Inducing resistance to conspiracy theory propaganda: Testing inoculation and metainoculation strategies. *Human Communication Research*, 39(2), 184–207. https://doi.org/10.1111/hcre.12000
- Basol, M., Roozenbeek, J., Berriche, M., Uenal, F., McClanahan, W. P., & Linden, S. V. D. (2021). Towards psychological herd immunity: Cross-cultural evidence for two prebunking interventions against COVID-19 misinformation. *Big Data & Society, 8* (1), 20539517211013868. doi:10.1177%2F20539517211013868.
- Bertolotti, M., & Catellani, P. (2018). The effects of counterfactual attacks on the morality and leadership of different professionals. *Social Psychology*, 49(3), 154–167. https://doi.org/10.1027/1864-9335/a000338
- Bertolotti, M., & Catellani, P. (2021). Hindsight bias and electoral outcomes: Satisfaction counts more than winner-loser status. *Social Cognition*, 39(2), 201–224. https://doi. org/10.1521/soco.2021.39.2.201
- Bode, L., & Vraga, E. K. (2018). See something, say something: Correction of global health misinformation on social media. *Health Communication*, 33(9), 1131–1140.
- Brehm, S. S., & Brehm, J. W. (2013). Psychological reactance: A theory of freedom and control. Academic Press.
- Bronstein, M. V., Pennycook, G., Bear, A., Rand, D. G., & Cannon, T. D. (2019). Belief in fake news is associated with delusionality, dogmatism, religious fundamentalism, and reduced analytic thinking. *Journal of Applied Research in Memory and Cognition*, 8 (1), 108–117. https://doi.org/10.1016/j.jarmac.2018.09.005
- Bruder, M., Haffke, P., Neave, N., Nouripanah, N., & Imhoff, R. (2013). Measring individual differences in generic beliefs in conspiracy theories across cultures: Conspiracy Mentality Questionnaire. *Frontiers in Psychology*, 4, 225. https://doi.org/ 10.3389/fpsyg.2013.00225
- Byrne, R. M. (2002). Mental models and counterfactual thoughts about what might have been. Trends in Cognitive Sciences, 6(10), 426–431. https://doi.org/10.1016/S1364-6613(02)01974-5
- Catellani, P., & Bertolotti, M. (2014). The effects of counterfactual attacks on social judgments. *Social Psychology*, 45(5), 371–381. https://doi.org/10.1027/1864-9335/ a000195
- Catellani, P., Bertolotti, M., & Covelli, V. (2013). Counterfactual communication in politics: Features and effects on voters. In I. Poggi, F. D'Errico, L. Vincze, & A. Vinciarelli (Eds.), *Multimodal communication in political speech. Shaping minds and social action* (pp. 75–85). Berlin: Springer Verlag.
- Catellani, P., Bertolotti, M., Vagni, M., & Pajardi, D. (2021). How expert witnesses' counterfactuals influence causal and responsibility attributions of mock jurors and expert judges. *Applied Cognitive Psychology*, 35(1), 3–17. https://doi.org/10.1002/ acp.3720

- Chan, M. P. S., Jones, C. R., Hall Jamieson, K., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science*, 28(11), 1531–1546. https://doi.org/10.1177/ 0956797617714579
- Compton, J. (2013). Inoculation theory. The SAGE Handbook of Persuasion: Developments in Theory and Practice, 2, 220–237.
- De Brigard, F., Addis, D. R., Ford, J. H., Schacter, D. L., & Giovanello, K. S. (2013). Remembering what could have happened: Neural correlates of episodic counterfactual thinking. *Neuropsychologia*, 51, 2401–2414. https://doi.org/10.1016/ j.neuropsychologia.2013.01.015
- De Brigard, F., Henne, P., & Stanley, M. L. (2021). Perceived similarity of imagined possible worlds affects judgments of counterfactual plausibility. *Cognition, 209*, Article 104574. https://doi.org/10.1016/j.cognition.2020.104574
- Douglas, K. M., Sutton, R. M., & Cichocka, A. (2017). The psychology of conspiracy theories. Current Directions in Psychological Science, 26(6), 538–542. doi:10.1177% 2F0963721417718261.
- Douglas, K. M., Uscinski, J. E., Sutton, R. M., Cichocka, A., Nefes, T., Ang, C. S., & Deravi, F. (2019). Understanding conspiracy theories. *Political Psychology*, 40, 3–35. https://doi.org/10.1111/pops.12568
- Ecker, U. K., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., ... Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1(1), 13–29. https://doi.org/ 10.1038/s44159-021-00006-y
- Effron, D. A. (2018). It could have been true: How counterfactual thoughts reduce condemnation of falsehoods and increase political polarization. *Personality and Social Psychology Bulletin*, 44(5), 729–745. doi:10.1177%2F0146167217746152.
- Effron, D. A., & Raj, M. (2020). Misinformation and morality: Encountering fake-news headlines makes them seem less unethical to publish and share. *Psychological Science*, 31(1), 75–87. doi:10.1177%2F0956797619887896.
- Epstude, K., & Roese, N. J. (2008). The functional theory of counterfactual thinking. Personality and Social Psychology Review, 12, 168–192. https://doi.org/10.1177/ 1088868308316091
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191. https://doi.org/10.3758/bf03193146
- Fiedler, K., & Mata, A. (2013). The art of exerting verbal influence through powerful lexical stimuli. In J. P. Forgas, O. Vincze, & J. László (Eds.), Social cognition and communication (pp. 43–61). Psychology Press.
- Forgas, J. P. (1995). Mood and judgment: The affect infusion model (AIM). Psychological Bulletin, 117, 39–66. https://doi.org/10.1037/0033-2909.117.1.39
- Galinsky, A. D., Liljenquist, K. A., Kray, L. J., & Roese, N. J. (2005). Finding meaning from mutability: Making sense and deriving significance through counterfactual thinking. In D. R. Mandel, D. J. Hilton, & P. Catellani (Eds.), *The psychology of counterfactual thinking* (pp. 110–125). Routledge.
- Greifeneder, R., Bless, H., & Pham, M. T. (2011). When do people rely on affective and cognitive feelings in judgment? A review. *Personality and Social Psychology Review*, 15, 107–141. https://doi.org/10.1177/1088868310367640
- Hayes, A. F. (2018). Partial, conditional, and moderated moderated mediation: Quantification, inference, and interpretation. *Communication Monographs*, 85(1), 4–40. https://doi.org/10.1080/03637751.2017.1352100
- Helgason, B. A., & Effron, D. A. (2022). It might become true: How prefactual thinking licenses dishonesty. Journal of Personality and Social Psychology. https://doi.org/ 10.1037/pspa0000308
- Hornsey, M. J., & Imani, A. (2004). Criticising groups from the inside and the outside: An identity perspective on the intergroup sensitivity effect. *Personality and Social Psychology Bulletin*, 30, 365–383. doi:10.1177%2F0146167203261295.
- Imhoff, R., & Bruder, M. (2014). Speaking (un-) truth to power: Conspiracy mentality as a generalised political attitude. *European Journal of Personality*, 28(1), 25–43. https:// doi.org/10.1002/per.1930
- Imhoff, R., & Erb, H. P. (2009). What motivates nonconformity? Uniqueness seeking blocks majority influence. *Personality and Social Psychology Bulletin*, 35(3), 309–320. doi:10.1177%2F0146167208328166.
- Imhoff, R., & Lamberty, P. K. (2017). Too special to be duped: Need for uniqueness motivates conspiracy beliefs. *European Journal of Social Psychology*, 47(6), 724–734. https://doi.org/10.1002/ejsp.2265
- Jolley, D., & Douglas, K. M. (2017). Prevention is better than cure: Addressing antivaccine conspiracy theories. *Journal of Applied Social Psychology*, 47(8), 459–469. https://doi.org/10.1111/jasp.12453
- Lee, A. M., & Chyi, H. I. (2014). Motivational consumption model: Exploring the psychological structure of news use. *Journalism and Mass Communication Quarterly*, 91(4), 706–724. doi:10.1177%2F1077699014550088.
- Lewandowsky, S., Cook, J., Oberauer, K., Brophy, S., Lloyd, E. A., & Marriott, M. (2015). Recurrent fury: Conspiratorial discourse in the blogosphere triggered by research on the role of conspiracist ideation in climate denial. *Journal of Social and Political Psychology*, *3*, 142–178. https://doi.org/10.5964/jspp.v3i1.443
- Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13(3), 106–131. doi:10.1177% 2F1529100612451018.
- Lewandowsky, S., Lloyd, E. A., & Brophy, S. (2018). When THUNCing trumps thinking: What distant alternative worlds can tell us about the real world. *Argumenta*, 3(2), 217–231. https://doi.org/10.23811/52.arg2017.lew.llo.bro
- Lewandowsky, S., & Van Der Linden, S. (2021). Countering misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology*, 32 (2), 348–384. https://doi.org/10.1080/10463283.2021.1876983

- Mancosu, M., Ladini, R., & Vassallo, S. (2021). Political consequences of conspiratorial thinking: Evidence from 2016 Italian constitutional referendum. Acta Politica, 56(1), 69–88. https://doi.org/10.1057/s41269-019-00146-3
- Markman, K. D., Lindberg, M. J., Kray, L. J., & Galinsky, A. D. (2007). Implications of counterfactual structure for creative generation and analytical problem solving. *Personality and Social Psychology Bulletin*, 33(3), 312–324. doi:10.1177% 2F0146167206296106.
- Markman, K. D., & McMullen, M. N. (2003). A reflection and evaluation model of comparative thinking. *Personality and Social Psychology Review*, 7, 244–267. https:// doi.org/10.1207/s15327957pspr0703_04
- Markman, K. D., & McMullen, M. N. (2005). Reflective and evaluative modes of mental simulation. In D. R. Mandel, D. J. Hilton, & P. Catellani (Eds.), *The psychology of counterfactual thinking* (pp. 77–93). Abington, UK: Routledge.
- Martel, C., Pennycook, G., & Rand, D. G. (2020). Reliance on emotion promotes belief in fake news. Cognitive Research: Principles and Implications, 5(1), 1–20. https://doi.org/ 10.1186/s41235-020-00252-3

McGuire, W. J. (1970). A vaccine for brainwash. Psychology Today, 36-39, 63-64.

McGuire, W. J., & Papageorgis, D. (1961). The relative efficacy of various types of prior belief-defense in producing immunity against persuasion. *Journal of Abnormal and Social Psychology*, 62, 327–337. https://psycnet.apa.org/doi/10.1037/h0042026.
Moscovici, S. (2020). Reflections on the popularity of 'Conspiracy Mentalities'.

- International Review of Social Psychology, 33(1). https://doi.org/10.5334/irsp.432 Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political
- misperceptions. Political Behavior, 32, 303–330. https://doi.org/10.1007/s11109-010-9112-2

Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., & Bang Petersen, M. (2021). Partisan polarization is the primary psychological motivation behind political fake news sharing on Twitter. *American Political Science Review*, 115(3), 999–1015. https://doi.org/10.1017/S0003055421000290

Pennycook, G., & Rand, D. G. (2020). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality*, 88(2), 185–200. https://doi.org/10.1111/jopy.12476

Petrocelli, J. V., Seta, C. E., & Seta, J. J. (2013). Dysfunctional counterfactual thinking: When simulating alternatives to reality impedes experiential learning. *Thinking & Reasoning*, 19(2), 205–230. https://doi.org/10.1080/13546783.2013.775073

Rabinovich, A., & Morton, T. A. (2010). Who says we are bad people? The impact of criticism source and attributional content on responses to group-based criticism. *Personality and Social Psychology Bulletin, 36*(4), 524–536. doi:10.1177% 2F0146167210362980.

Revlin, R., Cate, C. L., & Rouss, T. S. (2001). Reasoning counterfactually: Combining and rending. *Memory and Cognition*, 29, 1196–1208. https://doi.org/10.3758/ BF03206389

Roese, N. J. (1997). Counterfactual thinking. Psychological Bulletin, 121(1), 133. https:// psycnet.apa.org/doi/10.1037/0033-2909.121.1.133.

- Roese, N. J., & Epstude, K. (2017). The functional theory of counterfactual thinking: New evidence, new challenges, new insights. *Advances in Experimental Social Psychology*, 56, 1–79. https://doi.org/10.1016/bs.aesp.2017.02.001
- Roozenbeek, J., & Van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, 5(1), 1–10. https://doi.org/10.1057/s41599-019-0279-9

- Ross, R. M., Rand, D. G., & Pennycook, G. (2021). Beyond "fake news": Analytic thinking and the detection of false and hyperpartisan news headlines. *Judgment and Decision making*, 16(2), 484–504. https://doi.org/10.31234/osf.io/cgsx6
- Shalvi, S., Dana, J., Handgraaf, M. J., & De Dreu, C. K. (2011). Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. Organizational Behavior and Human Decision Processes, 115(2), 181–190. https://doi. org/10.1016/j.obhdp.2011.02.001
- Shen, L., & Dillard, J. P. (2005). Psychometric properties of the Hong Psychological Reactance Scale. Journal of Personality Assessment, 85(1), 74–81. https://doi.org/ 10.1207/s15327752jpa8501 07
- Sindermann, C., Cooper, A., & Montag, C. (2020). A short review on susceptibility to falling for fake political news. *Current Opinion in Psychology*, 36, 44–48. https://doi. org/10.1016/j.copsyc.2020.03.014

Starbird, K., Spiro, E., Edwards, I., Zhou, K., Maddock, J., & Narasimhan, S. (2016, May). Could this be true? I think so! Expressed uncertainty in online rumoring. In Proceedings of the 2016 CHI conference on human factors in computing systems (pp. 360–371). New York: ACM.

- Swami, V., Barron, D., Weis, L., Voracek, M., Stieger, S., & Furnham, A. (2017). An examination of the factorial and convergent validity of four measures of conspiracist ideation, with recommendations for researchers. *PLoS One*, *12*(2), Article e0172617. https://doi.org/10.1371/journal.pone.0172617
- Swami, V., Voracek, M., Stieger, S., Tran, U. S., & Furnham, A. (2014). Analytic thinking reduces belief in conspiracy theories. *Cognition*, 133(3), 572–585. https://doi.org/ 10.1016/j.cognition.2014.08.006
- Swire-Thompson, B., DeGutis, J., & Lazer, D. (2020). Searching for the backfire effect: Measurement and design considerations. *Journal of Applied Research in Memory and Cognition*, 9(3), 286–299. https://doi.org/10.1016/j.jarmac.2020.06.006
- Tal-Or, N., Boninger, D. S., Poran, A., & Gleicher, F. (2004). Counterfactual thinking as a mechanism in narrative persuasion. *Human Communication Research*, 30(3), 301–328. https://doi.org/10.1111/j.1468-2958.2004.tb00734.x

Tappin, B. M., Van Der Leer, L., & McKay, R. T. (2017). The heart trumps the head: Desirability bias in political belief revision. *Journal of Experimental Psychology: General*, 146(8), 1143–1149. https://psycnet.apa.org/doi/10.1037/xge0000298.

Uscinski, J. (2018). Down the rabbit hole we go! In J. Uscinski (Ed.), Conspiracy theories and the people who believe them (pp. 1–32). Oxford University Press.

Van Prooijen, J. (2019). Belief in conspiracy theories: Gullibility or rational skepticism? In J. Forgas, & R. Baumeister (Eds.), The social psychology of gullibility: Fake news, conspiracy theories and irrational beliefs (pp. 319–332). Taylor & Francis.

- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146–1151. https://doi.org/10.1126/science.aap9559
- Walter, N., & Tukachinsky, R. (2020). A meta-analytic examination of the continued influence of misinformation in the face of correction: How powerful is it, why does it happen, and how to stop it? *Communication Research*, 47(2), 155–177. doi:10.1177% 2F0093650219854600.
- Wong, E. M. (2010). It could have been better: The effects of counterfactual communication on impression formation. *European Journal of Social Psychology*, 40, 1251–1260. https://doi.org/10.1002/ejsp.719
- Wood, T., & Porter, E. (2019). The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behavior*, 41(1), 135–163. https://doi.org/10.1007/ s11109-018-9443-y