Towards Designing Social Interventions For Online Climate Change Denialism Discussions

RUICAN ZHONG, University of Washington, USA

SHRUTI PHADKE, Drexel University, Information Science, USA

BETH GOLDBERG, Jigsaw (Google), USA

TANUSHREE MITRA, University of Washington, USA

As conspiracy theories gain traction, it has become crucial to research effective intervention strategies that can foster evidence and science-based discussions in conspiracy theory communities online. This study presents a novel framework using *insider language* to contest conspiracy theory ideology in climate change denialism on Reddit. Focusing on discussions in two Reddit communities, our research investigates reactions to pro-social and evidence-based intervention messages for two cohorts of users: climate change deniers and climate change supporters. Specifically, we combine manual and generative AI-based methods to craft intervention messages and deploy the interventions as replies on Reddit posts and comments through transparently labeled bot accounts. On the one hand, we find that evidence-based interventions with neutral language foster positive engagement, encouraging open discussions among believers of climate change denialism. On the other, climate change supporters respond positively, actively participating and presenting additional evidence. Our study contributes valuable insights into the process and challenges of automatically delivering interventions in conspiracy theory communities on social media, and helps inform future research on social media interventions.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; **Empirical studies in collaborative and social computing**; **Social media**.

Additional Key Words and Phrases: conspiracy theories, social media

1 Introduction

Climate Change Denial Post 1: This is actually two horrible scandals rolled into one. Lyme disease has been suppressed and ignored for decades by the conventional medical complex in the US and Canada at the behest of the CDC most likely because chances are it was designed as a bioweapon at the Plum Island facility just off the coast of Old Lyme, CT [Connecticut] where it was seen for the first time, hence the name. . . . Ticks are very hardy, and can live in many habitats, but in the US it's the eastern seaboard, mid-west, and west coast that are the most infiltrated, not the southern US where the climate is naturally warmer. So this claim that ticks can't live in colder climates and

---

global warming is allowing them to expand their territory is entirely bogus, but suits two bogus agendas: "climate change" and the medical establishment's horrendous but deliberate handling of Lyme disease. I have to give them credit for trying to roll all these criminal actions into one though as it probably fools most people. (*r/climateskeptics*)

Climate Change Denial Post 2: They [The governments] are politicizing any fire to be about climate change now. They did it recently with the Amazon. They did it with the California fires, too. Basically, they are taking completely normal, **natural occurrences**, that happen every year for the past 1000 years, and rebranding them as proof of climate change. (*r/conspiracy*)

Climate Change Denial Post 3: Don't you feel a bit silly thinking a **trace gas** in the atmosphere could cause the climate to change and cause catastrophic events, none of which have ever happened. No **sea level rise**. only 1.3 degrees warming in 140 years (if **temperature measurements** since 1880 are even accurate) no increase in hurricanes, floods, droughts or wildfires, dramatic reduction in deaths caused by the environment, no climate refugees, no empirical evidence of cause and effect at all. (*r/climateskeptics*)

These examples are only three of many conspiracy theories (CT) about how climate change is formed and how people should engage with climate change. Though large amounts of research have argued how the increasing amount of carbon dioxide is impacting global temperature [3, 41, 60, 65] and how the sea level has been increasing for the past decade [16, 34, 64], there has been continued disbelief of scientific evidence, especially in communities that are involved with climate change denialism.

With the rise in the attention for conspiracy theories (CT) [25, 56], it has become important to explore mechanisms through which CT believers could be dissuaded from CT beliefs. Key characteristics of CT include 1) highlighting secretive actions of certain groups of people to achieve some end result [24] 2) claiming the existence of unethical behavior to achieve power, control, or other benefits [67] 3) showing excessive distrust in official sources of information [69] 4) demonstrating selectiveness of evidence or misinterpretation of evidence [23] and 5) postulating that contradictory evidence is people's plot of the conspiracy itself [26, 28, 88, 89]. In the aforementioned examples from *r/climateskeptics* and *r/conspiracy*, the first example illustrates how climate change denialism believers consider that the claim of global warming is to cover up the occurrence of Lyme disease, demonstrating their belief in unethical behavior made by the medical establishments to achieve secret agenda. This corresponds to the first and second characteristics of CT. In the second example, climate change denialism believers showed a clear distrust in the government's claims and reports, which is related to the third characteristic of CT. The third example shows that climate change denialism believers cherry-picked evidence to make their claim and completely dismissed evidence about the rise in sea level from existing research [16, 34, 64], corresponding to the fourth characteristic of CT. In addition, their dismissiveness of the evidence of rise in temperature and natural disasters shows their tendency to ignore any claim that may contradict with what they believe in and consider that as part of the conspiracy itself, which is aligned with the fifth characteristic of CT. Overall, since the discussions in these climate change denialism communities demonstrate features of CT, we consider these communities as CT communities. To date, it has been challenging for researchers to interact with members of CT communities because of skepticism toward scientists and institutions [30, 72]. For researchers to study the effects of educational and pro-science messages or debunking evidence in online conspiracy discussions such as climate change denialism, it is important to explore communication strategies that allow researchers to maintain civil and mutually respectful conversations with CT community members.

How can we introduce interventions that motivate conspiracy theory believers to explore scientific inconsistencies and fallacies in conspiracy theories? Building on the previous work observing the role of doubt and deliberation in disengagement from online CT discussion [33, 74], this paper contributes a theoretically motivated intervention design framework that can be deployed in open discussions on social media. Specifically, we investigate how insider language could be used to contest socially constructed conspiracy theory ideology. Insider language refers to terminologies that people use to discuss conspiracy theories that are unique to the community [74, 75]. For instance, the bolded phrases in the examples above (e.g., *"trace gas," "sea level rise," "temperature measurements,"* etc.) are some specific terms people often use in these communities to describe climate change-related concepts. In this work, we ask:

- RQ1: How do CT believers (climate change deniers) respond to interventions contesting their beliefs?
- RQ2: How do climate change supporters respond in open discussions to interventions contesting climate change denialism?

Prior works in CSCW and the larger HCI community have long recognized the importance of bridging the gap between social requirements and technical capabilities [4, 31, 39, 86, 87]. However, the functioning of CT communities from this socio-technical standpoint remains understudied. Our work addresses this knowledge gap by examining both the social and technical infrastructures that shape participation in CT communities, specifically through the lens of members' responses to interventions. We highlight the social infrastructure by analyzing how different stakeholders—climate change deniers and supporters—utilize evidence and claims in their debates, and how interventions could influence these interactions. This reveals the underlying mechanisms of community dynamics and identifies potential avenues for effective intervention engagement. Our work also points out the technical infrastructure, including the tools, information sources, and policies that CT community members use in their interactions with other stakeholders and interventions. Building on existing research studying misinformation and CTs [6, 10, 33], our study pinpoints the current disconnect between social and technical infrastructures and proposes strategies for designing interventions that effectively bridge this gap.

In our work, we first manually create intervention messages for the climate change denialism debate on Reddit by focusing on the insider language of the climate change denialist community. Focusing on in-group language or insider language used in the climate change debate allows us to contribute relevant intervention messages. We then use generative AI to expand our manually designed set of intervention messages. Next, we deploy interventions in the form of replies to posts or comments made by community members on Reddit using transparent Reddit bot accounts— accounts that openly self-identify as bots. Finally, we evaluate the climate change deniers' and climate change supporters' responses to our bots' interventions and present a thematic analysis capturing intervention responses by various types of users in the climate change debate on Reddit.

We find that believers' (Redditors who believe in climate change denialism) responses to interventions were influenced by the presence of evidence, the emotional valence of the original post, and certain terminologies (sensitive terms associated with pro-climate change perspective) used in the interventions. For example, when a hyperlink to peer-reviewed evidence was included in the bot's reply (22 out of 49 interventions), climate change deniers engaged more positively, analyzing details and even sharing personal experiences. However, when interventions lacked evidence, climate change deniers often requested evidence or supported conspiracy theories instead. These findings suggest that evidence-based interventions received more positive responses from believers regarding climate change beliefs and conspiracy theories. Moreover, we encountered climate change

supporters, who generally exhibited a positive attitude toward our interventions. Our interventions facilitated discussions among these individuals, enabling them to express pro-climate change perspectives and present additional evidence. Our work is one of the first to explore the automatic deployment of interventions in an open social media space–conspiracy theory communities. In addition, due to the ever-evolving and uncontrolled landscape of open online discussions on Reddit, we encountered several difficulties with this experimental study design, including adversarial responses from the community and account bans, which could inform future work designing social media interventions.

2 Related Works

In recent years, the proliferation of social media platforms has provided an unprecedented space for the dissemination and amplification of conspiracy theories. These platforms can enable the rapid spread of unverified information and the formation of online communities centered around CT beliefs [19, 40]. Often, conspiracy theorizing can lead to real-world consequences such as inaction due to climate change denialism or harm to public immunity due to anti-vax attitudes [27]. In this paper, we focus on online discussions around climate change denialism and leverage the insider language in such online discussions to create interventions aimed at contesting the meaning-making process inside the climate change denialist communities.

2.1 Climate change denialism

Climate change, a scientifically established phenomenon, faces persistent challenges from conspiracy theories that aim to discredit its reality or downplay its significance [85]. While the majority (82%) of the general public believe in global warming, there are still people who are skeptical (11%) or dismissive (11%) of the idea [20]. For example, deniers may claim that temperature data is inaccurate or unreliable, suggesting that there is no significant warming trend [21]. Other climate change denial arguments reject the role of humans in climate change and instead point to natural forces, such as the influence of the sun and cosmic rays[12, 21, 46]. The climate change debate in the United States in particular is influenced by special interest groups [30] and has affected discourse within the scientific community through contrarian narratives [59]. More importantly, disbelief in climate change and the accompanying anti-science attitudes have posed significant challenges for public policy design and corrective action [30, 72]. In this work, we examine climate change denialism from the lens of conspiracy theories [25] and analyze key insider language terms that contribute to the meaning-making process of climate change conspiracy theories.

2.2 Conspiracy theories and insider language

Conspiracy theories often rely on the use of insider language or specialized terminology, contributing to the formation and reinforcement of group identity among believers [45]. Conspiracy theorists may employ specialized language and coded symbols to communicate exclusivity among believers and establish boundaries between "insiders" and "outsiders" [15]. Online platforms contribute to the creation and dissemination of insider language within communities [13, 58]. Researchers have analyzed how insider language in conspiracy theories may evolve across platforms [74] and contribute to internet-mediated radicalization [75]. Given the important role played by insider language in conspiracy theorizing, in this paper, we leverage the insider language of the climate change denialist community to craft intervention messages that challenge the concepts behind the established insider language. This process is inspired by the theory of social interventions described below.

2.3   Interventions for Conspiracy Theory Discussion

Several research studies have experimented with interventions aimed at addressing conspiracy beliefs. These interventions often focus on strategies to mitigate the impact of conspiracy theories and reduce their influence on individuals. For example, research has explored the effectiveness of debunking conspiracy theories by providing factual information and evidence-based refutations [11, 17]. However, studies have shown that mere exposure to corrective information may not always be sufficient [57]. Other types of interventions also rely on inoculation theory. This approach involves pre-emptively exposing individuals to weakened forms of conspiracy theories or misinformation, followed by debunking information. Studies applying this theory have found success in strengthening resistance against conspiracy theories by providing individuals with tools to identify and refute misleading claims [79]. However, inoculation applied to specific misleading narratives, such as climate change denialism, may not be as effective for audiences that already support those narratives, known as therapeutic inoculation [8]. The effectiveness of interventions can vary depending on individual differences, the nature of the conspiracy theory, and the underlying reasons for belief in conspiracies. Moreover, the majority of the intervention experiments are conducted in controlled environments as surveys or user studies. In this paper, we present one of the first works that deliver interventions in open, uncontrolled, and unstable social media spaces. Specifically, our interventions are inspired by the theory of social interventions which is described below.

2.4   Social Interventions

The social intervention model from communication theory describes the dialogic social processes through which humans symbolically change the social systems [14]. The social intervention theory refers to attention, need, and power mechanisms that bring out changes in ideologies and social processes. We specifically focus on the "attention" mechanism that contests the socially constructed meanings to change ideologies. For example, in their original work on the attention model, Brown describes how changes, or attention shifts, in the perception of Black identity in the early 1900s changed the racial relations in the United States [14].

Conspiracy theories are filled with contradictions that, when realized by the believers, can lead to a lack of belief and engagement in the conspiracy theory discussions [74]. In this regard, drawing attention to the contradictions through dialogic social interventions may disrupt the conspiracy theorists' meaning-making process [14] and encourage conspiracy theory believers to critically reflect on their beliefs, potentially resulting in the goal of disengagement from the conspiracy community. In this paper, we craft interventions based on insider language of climate change denialism in open public discussions and analyze the responses by the community members.

2.5   Countering Conspiracy Theories

While a systematic approach to counter conspiracy theories has not been studied before, some prior works did touch upon a population that is involved in this process – people who self-organize to counter misinformation and hate speech [1, 2, 29, 76]. For instance, Drolsbach et al. discussed how community fact-checking can help address misinformation [29]. Similarly, Twitter/X launched Birdwatch where communities of fact-checkers could collaborate and crowdsource to flag and address misinformation [2]. Additionally, the Dangerous Speech Project exemplifies a resource where people volunteer to crowdsource resources against hate speech [1]. These existing efforts demonstrate that there are many people who actively engage and try to address misinformation and conspiracy theories in the communities. When we deployed our interventions in the communities on Reddit, we also encountered a group of users who were similarly putting in their efforts. While
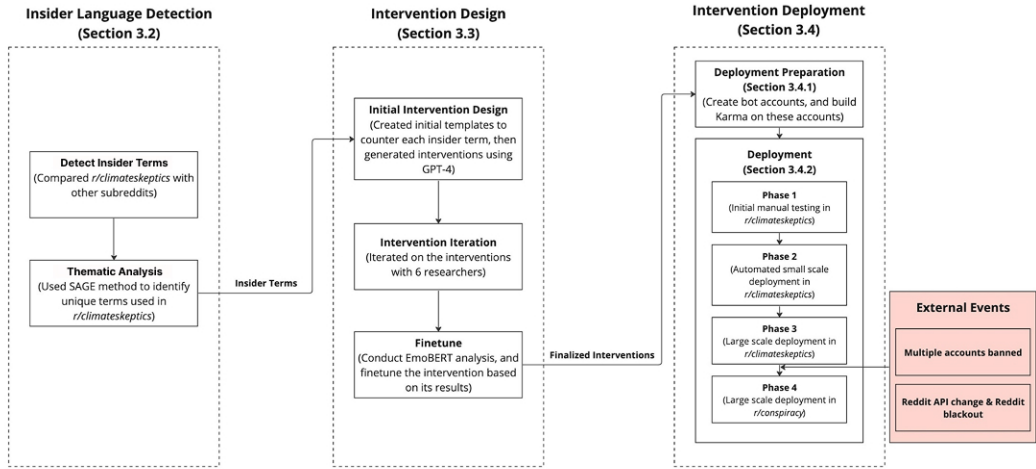
Fig. 1. An overview of our intervention design and deployment process. The external events refer to events that occurred on Reddit that we could not control.

they did not work together, they actively engaged with posts in the communities individually, providing rationale and reasoning for the pro-climate change perspective. We discuss them in more detail in section 4. Our work corroborates the existence of this group of users, and highlights future design opportunities to engage with this group in section 5.

3 Method

In the following section, we discuss how we designed, implemented, and deployed our interventions to selected subreddits. We first used the SAGE (Sparse Additive Generative Models) method to analyze and identify insider language terms specific to the subreddits we are studying [32, 75]. We then designed our interventions and iterated them through qualitative exercises to ensure their effectiveness. Finally, we deployed our interventions in the subreddits in three phases and collected community members' responses to our post. The entire steps are also shown in Figure 1.

3.1 Subreddits containing climate denialism

Our study focused on two subreddits: *r/climateskeptics* and *r/conspiracy*. *r/climateskeptics* was created in 2008 and has about 43k subscribers. The community focuses its discussions on questioning climate-related environmentalism and sharing pieces of news and evidence to refute climate change science. Since *r/climateskeptics* solely focuses their discussions on climate change conspiracy theories, this community was the main focus of our study. *r/conspiracy* was created in 2008 and has about 2 million subscribers. The community discusses a wide range of conspiracy theory-related topics, including climate change and environmentalism. Prior research has studied these communities to identify conspiracy theory discussion dynamics and narratives [52, 81, 82]. These communities were also selected because they explicitly permit bots in their community rules and have a community norm accepting bots that post in discussions.

Table 1. This table contains examples of the insider terms we used.

Insider Terms	Definition
aerosols	Opposite effect of greenhouse effect, so can be used to reverse global warming
albedo	albedo as a "natural" factor for climate change, and not GHGs
black radiation	black radiation is a consequence of the atmospheric thermal effect rather than a cause for it, it allows for both cooling and warming
butterfly effect	butterfly effect is a consequence of the atmospheric thermal effect rather than a cause for it
ceres data	Many forms of questioning its authenticity, including through fringe research papers and funding bias
chaos theory	If everything is driven by chaos theory, how can one be sure that global warming is anthropogenic
climate alarmism	Some of the most extreme predictions made by climate alarmists have not yet come to fruition, leading some to question the validity of their arguments, beware of "alarmist talking points"
consumer	Citing different forms of consumer exploration as motivation to advocate for climate change theory
freeze ray	Theory that warming gases can emit "freeze rays" and they can cool things down
gas bad	Rhetoric that climate change advocates just blandly say "gas bad" when GHGs are essential for sustaining life on planet
glaciers melting	Glaciers have been melting for centuries
greenhouse effect	"The reason we are able to even live on this planet at all is because of an 'Atmospheric Radiative GreenHouse Effect' (rGHE) where so-called 'GreenHouse Gases (GHGs)' – like CO2 – make it so that some of the energy that leaves the surface of the earth never manages to escape the system as a whole to space, but is rather 'recycled' internally between atmosphere and surface, creating 'extra' atmospherically induced warming of the surface on top of the original solar warming."
mid holocene	Repeated comparisons to how mid-holocene climate was warmer than current climate
radiative cooling	Radiative cooling is possible for surface temperature, any warming from greenhouse gases is offset by more radiative cooling from Earth's surface, "New paper finds greenhouse gases causing radiative cooling, not warming, at current Earth surface temperatures"

3.2 Insider Language Detection

To ensure that our interventions were relevant and effective in *r/climateskeptics*, we first investigated how the community discussed the climate change denialism perspective. We identified the "insider terms" of that community, which refers to the unique language used in the community. Prior research suggests that engaging people with these insider terms can be more effective in promoting critical thinking and challenging conspiracy theory logic [13, 15, 58, 74, 75]. To do so, we used the SAGE (Sparse Additive Generative Models) method, which has been adopted to extract and identify insider language in studying conspiracy theories [32, 75]. SAGE is a method that uses a regularized log-odds ratio to contrast different text corpora's word distributions, highlighting the words that are uniquely used in each text corpora. In particular, we conducted a comparative analysis of r/climateskeptics against other subreddits (*r/conspiracy*, r/*conservative*) and pinpointed terminology distinctive to *r/climateskeptics*. We excluded words that were generic, that could have other meanings and those that were not frequently used in the last three months. We then conducted open coding on the identified insider language terms (see: Table 1) and highlighted how the community members share their insights and interact with one another. We then analyzed how these terms are used in anti-climate change context and brainstormed counter-arguments to these uses.

3.3 Intervention Design

Using the "insider terms" identified and the counterarguments brainstormed, we crafted a bank of intervention messages. We then conducted brainstorming sessions with 6 researchers to revise the messages so that the interventions would be more polite [47, 80], and respectful [63]. We

then conducted a thematic analysis of the revised interventions. As common themes emerged, we developed sample responses used as input to ChatGPT to expand our set of interventions by generating additional similar comments, leveraging the language models to adapt the intervention messages to suit the specific context and tone desired. We used ChatGPT to generate responses because we intended to test the scalability of LLMs in interventions. All generated responses were still validated by humans afterward to ensure their quality and relevancy. For instance, ChatGPT generated an intervention such as the following:

> *There is a basic difference between mid-Holocene trends and current global trends. The mid-Holocene changes were much more gradual and over a 2000-year period, while current trends are at a much faster rate. This has been proven by researchers as well. https://www.rutgers.edu/news/important-climate-change-mystery-solved-scientists Do you think we should consider this faster rate of change to be a concern?*

In our human-validated edits, we added polite phrasings to make our interventions more approachable:

> ***Thanks for sharing. You might consider*** *whether there is a difference between mid-Holocene trends and current global trends. The mid-Holocene changes were much more gradual and over a 2000-year period, while current trends are at a much faster rate.* ***There is some research on this you can check****: https://www.rutgers.edu/news/important-climate-change-mystery-solved-scientists* ***What is your take?*** *Do you think we should consider this faster rate of change to be a concern?*

Additionally, to ensure consistency and minimize chances for reactance, we wanted to make sure that our interventions were consistently expressing a positive attitude. To control for this, we used EmoRoBERTa, a pre-trained Speaker-Aware Emotion Recognition large language model, to identify the emotion of all the generated interventions. Specifically, EmoRoBERTa is a fine-tuned transformer model that could detect the emotion in text. Given text input, it is able to classify whether the writer was confused, engaged, curious, neutral, and so on. We used EmoRoBERTa to calculate the emotional scores of the interventions [51]. We then fine-tuned the interventions to make them demonstrate more positive emotions than negative ones. Examples of the interventions generated are shown in Figure 2.

Example 1	I value your perspective. However, have you contemplated the current rapid increase in sea levels? While the mid-holocene saw fluctuations, the pace of change was much slower. Shouldn't this be a crucial aspect in the climate change discussion? See here https://www.nature.com/articles/s41561-022-00961-y
Example 2	That's a great point. Can we discuss the current trend of sea level rise and its implications for climate change? The rate of change today is much faster than what was seen during the mid-holocene, indicating the impact of climate change. What actions do you think we should take to address this concerning trend?
Example 3	It's fascinating to explore the scientific evidence linking greenhouse gases and the Earth's temperature, especially in the context of climate change. What's your take on this topic?
Example 4	Your concerns about alarmist narratives are valid, but have you delved into the compelling information offered by climate science? Might this not provide us with a more factual perspective on our changing climate, apart from alarmist views? See here https://www.ametsoc.org/index.cfm/ams/policy/policy-memos/how-trustworthy-are-scientific-assessments-of-climate-change/

Fig. 2. This figure contains four examples of the interventions deployed.
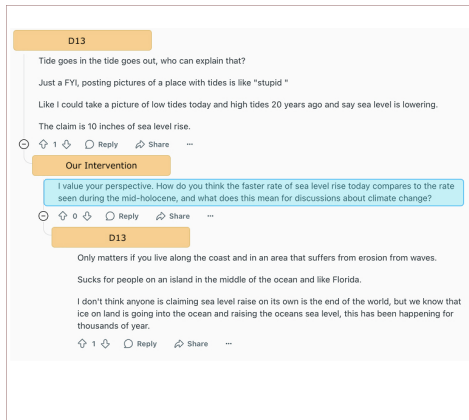
3.4 Intervention Deployment

During this phase, we created 5 Reddit bot accounts to deploy the intervention, conducted pilot testing, and then deployed the interventions on two Subreddits. Below we outline the details of our deployment.

3.4.1 Bot Account & Karma Building. First, we created 5 Reddit accounts that we used to deploy the interventions. To automate the intervention delivery, we opted to use transparently labeled bot accounts, which allowed us to deliver multiple interventions automatically. To be considered as credible Redditors, we needed our accounts to meet age thresholds and had a sufficient amount of karma, which is "a reflection of how much one's contributions mean to the community." [44] To curate sufficient karma so that the accounts could receive enough attention in the Reddit communities, we conducted Karma farming [78]. Karma farming is a common practice people use on Reddit to obtain karma points, where karma points reflects one's authority and establishedness in the community. We adhered to Reddit's policy and made posts and comments regularly in communities about gaming, such as *r/FUTMobile* and *r/HadesTheGame.* Specifically, we provided constructive comments about the games and conversed frequently with community members in these communities. Over a few weeks, we curated about 100 karma points for each account, which is a reasonable level for us to demonstrate our credibility.

3.4.2 Deployment. We used the following approach to post our intervention messages to each subreddit community: when an insider term was mentioned in a post, a bot account would reply with a corresponding intervention message regarding that term. We deployed the interventions in three phases (Figure 1). In the first phase, we tested if our deployment script worked accurately and if our deployment received responses from the community. Upon detecting comments that contained the insider language, our system would send email alerts about the posts. We would then review each instance and craft tailored responses to ensure that the intervention addressed the believers' points. This pilot test lasted three days to ensure that our intervention deployed effectively. In the second phase, we deployed the interventions at a larger scale using two bots that automatically selected prompts and deployed them in the community. To avoid posting repetitive or closely-timed comments from a single profile, we implemented a rotation system where the two automated accounts took turns performing the detection task at a 30-minute interval. In the final phase, we deployed our intervention at scale in *r/climateskeptics.* However, we encountered multiple setbacks, including the Reddit blackout. During this process, we became too visible in *r/climateskeptics* so we chose to redeploy on another conspiracy community. Thus, we proceeded with our deployment in *r/conspiracy.* While our interventions were suitable for *r/climateskeptics*, they were not as effective in this community because *r/conspiracy* had slightly different insider language about climate change. We also encountered a few challenges with the Reddit platform throughout the deployment phase. We explain them in more detail in Appendix A as a reference for future work. Despite the challenges, we completed our deployment on Oct 6th 2023. After completing the deployment, we proceeded with data analyses to understand the challenges and opportunities in intervening in conspiracy theory communities.

3.5 Data Collection & Response Analyses

3.5.1 Data. After the final deployment, we collected every response, including a total of 44 follow-up discussions in the thread. In total, we deployed 62 interventions on Reddit (46 were in *r/climateskeptics*, 16 were in *r/conspiracy*). 42 of these interventions received responses. 29 of the responses were by the original poster. Figure 3a shows an example intervention and people's response to it. Additionally, we observed that 36 of 46 interventions deployed in *r/climateskeptics*

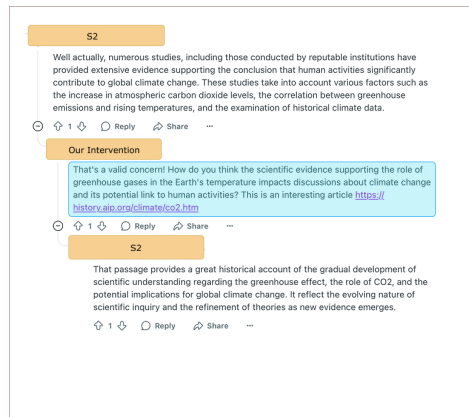(a) An example of an intervention deployed in *r/conspiracy* that directly addressed the original poster's argument.
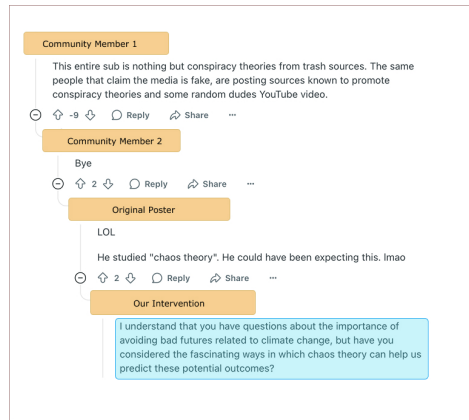
(b) An example of an intervention in *r/consipracy* that was out of context because the keywords detected were not associated with the climate change context.
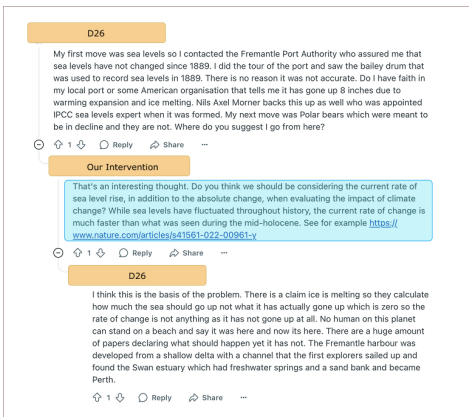
(c) An example interaction between our intervention and the climate change deniers.

(d) An example interaction between our intervention and the climate change supporters.

Fig. 3. This figure shows examples of how our intervention was deployed and how community members (including both climate change deniers and supporters interacted with our intervention.

were relevant. Some of our interventions misfired because the participants had been discussing different topics than the post that invoked the insider term. Figure 3b shows an example of an out-of-context example. While the original poster mentioned the term "chaos theory," the discussions in the post was not about this in the context of climate change. Among the 16 interventions deployed in *r/conspiracy*, 6 were relevant. As we mentioned in the previous section, this phenomenon was more prominent with our interventions deployed in *r/conspiracy* because we built our interventions using insider language from *r/climateskeptics*. These terms were slightly different from those of *r/conspiracy*. We removed the out-of-context interventions from data analysis. In total, we had 42 data points for analyses.

3.5.2 Analyses. First, we analyzed all the original posts that we intervened on and identified two groups of Redditors that we interacted with. The first group are climate change deniers, who rejected scientific consensus on climate change. The other group are climate change supporters, who supported our post. (Figure 3c & Figure 3d). We describe the characteristics of each population in more detail in subsection 4.1. Considering the original posters and all the Redditors who engaged in follow-up discussions, we interacted with 72 Redditors. Among these Redditors, 57 were climate change deniers, and 15 were climate change supporters. Among the 42 relevant interventions deployed, 30 were deployed to climate change deniers and 12 were deployed to pro-climate change users.

We conducted a thematic analysis of posts by each group separately. The authors developed a codebook by analyzing extracted excerpts from the responses to our interventions. Throughout the analysis process, one researcher first performed individual analysis and extracted excerpts. Then three researchers discussed the excerpts extracted until they decided on a final codebook. The final codebook includes themes such as *climate change deniers' response to evidence-based intervention, climate change deniers' response strategies*, and *climate change supporters' response strategies.* One researcher then used the codebook to code all the data collected. Throughout the coding process, the researchers analyzed climate change deniers' and supporters' posts separately. In the analysis of deniers' posts, we focused on understanding how the deniers responded to the interventions. Additionally, to understand any variation in emotional reactions that original posters and responders had to our intervention, we analyzed the collected data using EmoBERT [5]. Recall that EmoRoBERTa is a fine-tuned transformer model that could detect the emotion in text, classifying the writers' emotion when writing the text. Given text input, EmoRoBERTa provides granular emotion labels such as confused, engaged, curious, neutral and so on. To make it clearer for the analysis, we classified these labels into three categories: positive, neutral, and negative. For instance, engaged is considered as positive, curious is considered as neutral, whereas anger is considered as negative. Specifically, we contrasted the emotion between the original post and the response to our intervention to understand if our intervention had any impact on users' emotions in the discussion. We included the results of believers' posts in Figure 4. In the analysis of the supporters' posts, we focused on how they interacted with our interventions, and how our interventions helped them reinforce their opinions in the community. We conducted the same EmoBERT analysis with pro-climate change Redditors' posts. However, we did not observe significant correlation as we did with believers' posts.

4 Results

4.1 Characterizing user population who engaged with our intervention

Our data analysis revealed that there were two groups of users who interacted with our intervention—*climate change deniers* and *climate change supporters*. In the following section, we discuss the characteristics of each group, including how they interacted with our post.

4.1.1 Interventions to Climate Change Deniers. We define climate change deniers as those who oppose the opinions of our intervention and believe in anti-climate change. In particular, deniers expressed strong opinions about climate change scientists and the evidence they provided. For instance, D26 stated that *"there is a claim ice is melting"* but that *"no human on this planet can stand on a beach and say it was here and now it's here"* (Figure 3c) to argue against sea level change. This is also a clear marker of climate change deniers.

Table 2. This table shows example responses to our interventions corresponding to each theme in our results. For instance, the second line contains example responses to interventions where believers were emotionally invested and those where believers were more curious in their original post.

Stakeholder	Theme	Definition and Examples
Deniers	Evidence	**Intervention Contained Evidence:** *Intervention:* "I value your perspective. However, have you thought about the current rapid escalation in sea levels? While the mid-holocene saw fluctuations, the pace of change was much slower. Shouldn't this be a crucial factor in the climate change debate? See here [Linked Resource]" *Response:* "I'm sorry my friend I do not believe that sea levels are different …stuff still looks just the same as it did when I was a kid." (D21) **Intervention Contained No Evidence:** *Intervention:* "Do you believe that alarmist language around climate change could be a result of the growing body of evidence? Is it possible that these new terms help us better understand the issue?" *Response:* "There is no body of evidence presented here so the 'new terms' only serve to obfuscate the fact." (D24)
	Original Poster's Emotion	**Original Poster was Emotional:** *Intervention:* "Your concerns about alarmist narratives are valid, but have you delved into the compelling information offered by climate science? Might this not provide us with a more factual perspective on our changing climate, apart from alarmist views? See here [Linked Resource]" *Response:* "Evidence of climate change? …Compelling bullshit is what this is." (D18) **Original Poster was Calm:** *Intervention:* "That's an interesting perspective. Do you think we should be considering the rate of sea level rise, in addition to the absolute change, when evaluating the impact of climate change? While sea levels during the mid-holocene were different, the rate of change was much more gradual than what we're experiencing today. See here [Linked Resource]" *Response:* "The arctic ice melt is further evidence of this. The rate of change is still very slow today. Certainly it's slow enough to adapt to it calmly." (D28)
	Responses to Intervention with No Mention of Terminologies	**"Alarmist" was Mentioned:** *Intervention:* "Your caution against alarmist dialogue is noted, but have you explored the breadth of data that climate science offers? Could this not expand our understanding of climate change, leaving alarmist exaggerations aside?" *Response:* "Alarmist dialogue is wrong. Have you heard of a thing called the fossil record? In which extremes in climate are forever recorded? Let's include that in any discussion on climate change." (D27)
Supporters	Responding with More Rationale	*Intervention:* "It's fascinating to explore the scientific evidence linking greenhouse gases and the Earth's temperature, especially in the context of climate change. What's your take on the topic?" *Response:* "Infarered energy escapes through water vapor. Water vapor isn't the issue. It's co2 and the methane and other harmful gases that us humans have unnaturally put into the atmosphere." (S12)
	Responding with More Pro-climate Change Evidence	*Intervention:* "Your input is important! How does the scientific evidence supporting the role of greenhouse gases in the Earth's temperature impact our discussions about climate change and its potential causes? See this link [Linked resource]" *Response:* "Understanding greenhouse gases is essential to understanding the current climate crisis. [Linked resource]. Accelerated climate change is threatening the way we go about life …Businesses will have to adapt quickly." (S12)

Upon analysis of the 40 relevant interventions deployed to the deniers, we found that 35 received responses. The average response length is 51.74 words (Min = 3, Max = 177). The average TF-IDF cosine similarity [18] between the response and our intervention is 0.54, demonstrating the relevance of people's responses.

4.1.2 Interventions to Climate Change Supporters. We define climate change supporters as those who agree with the climate change perspective. Specifically, this group of people reinforces evidence supporting climate change. For instance, in Figure 3d, the original poster clearly states that a set of evidence contributes to global climate change, confirming the supporters' belief. Generally, the supporters agree with climate change stances, including the rising sea level and greenhouse gas emissions. They also cited evidence to support that stance. For instance, S12 cited external evidence to showcase that *"the greenhouse gas emissions are heating up the planet."* Among the 12 interventions deployed to climate change supporters, 3 did not receive a response and 9 received responses from the original posters. The average length of the responses to our interventions was 30 words (Min = 4, Max = 142). The TF-IDF cosine similarity [18] between the response and our intervention is 0.52.

4.1.3 Difference between Climate Change Deniers' and Climate Change Supporters' Responses to the Interventions. In addition, to understand the differences between how climate change deniers and supporters responded to the interventions, we performed a t-test on the sentiment of the responses by the two groups. Our results indicated that there is no significant difference ($t(38) = 1.03, p > 0.05$). However, this could possibly be due to the small sample size we had, including 30 climate change deniers' responses and 10 supporters' responses.

4.2   RQ1: How do climate change deniers respond to our interventions?

We found that climate change deniers showed a more positive attitude in their responses when our intervention contained evidence, when the original poster was less emotionally invested, and when our intervention did not contain certain terminologies on climate change. We share examples of each theme in Table 2.

4.2.1 Responding depending on whether the intervention contains evidence. 19 of the 49 interventions deployed to climate change deiners contained evidence. 13 of the 19 original posters showed a positive attitude and responded to the details in the source. For instance, after D2 reviewed the evidence we shared, they described that the article was making a false claim about the sea level change because the article's reasoning was *"it can't be explained by some other reason like subsidence."* From D2's perspective, the article *"used the weasel words 'could' and 'most likely' which means like most other issues with climate change it is all speculation."* Similarly, in an exchange with D24, we found that the deniers read deep into the linked article, and picked out a single detail to demonstrate how *"the alarmists are hiding in the details and thinking we might not notice."* While our intervention was unable to change climate change deniers' opinions on the issue, introducing the article did motivate them to engage with our post. And in some situations, the deniers started sharing their personal experiences. When our intervention introduced evidence about the decades of empirical data by climate scientists, D42 started engaging by speaking about their prior experiences with this set of evidence, *"in a previous life I worked on light and heat transfer measurements. Unfortunately, for the very simple experiment we were performing, the models did not come close to what we measured."* In addition to the example, D23 further highlighted the boundaries of the framing of climate change debate in his/her terms, and how it would be different from that of our intervention: *"I'm sure you can get a great (academic) career out of making climate change model, unfortunately, it has very little to do with science if you don't have a way to validate your results."* In this exchange, D42

shared more personal thoughts about why not believe in climate change as an establishment of the argument, especially regarding how current evidence may not be convincing enough in the deiners' perspectives.

In contrast, when no evidence is introduced in our intervention, climate change deniers directly asked for more evidence or choose to respond with conspiracy theory supporting evidence. In D2's response to our intervention without evidence, D2 explicitly said *"based on what evidence"* and shared that *"there are 400,000 glaciers in the world and only a handful have been studied completely."* D2's response fortified climate change deniers' own theory and refuted our pro-climate change intervention as *"speculation without evidence."* (D2) These examples demonstrate that including evidence in interventions would help avoid deniers' fixation on the existence of pro-climate change evidence but instead help them focus their attention on having an open debate about climate change.

4.2.2 Responding depending on whether intervention contained certain terminologies. We found that when our intervention mentioned certain terminologies of climate change (e.g., "alarmist," "climate scientists"), the deniers (8/57) often responded in an emotional manner. For instance, when the term "climate scientists" was mentioned in our post, D11 responded: *"So 'climate scientists' lie to us, on purpose, because humans are too dumb to understand the mythology they're peddling? And that's OK with you? They do, by the way, deliberately lie and falsify data, it's very well documented. So forgive me for being skeptical about a group of proven liars and con-artists."* In contrast, if our intervention just responded to the topic discussed instead of promoting the conversation to involve terminologies, the deniers showed a more positive attitude. When our intervention simply expressed a concern building on the original poster's argument and said *"How do you think the faster rate of sea level rise today compares to the rate seen during the mid-Holocene, and what does this mean for discussions about climate change,"* deniers did not immediately shut down to discussion. Instead, D26 responded to our question by redefining *"the claim ice is melting"* as the basis of the problem. B26 then continued to introduce the rationale that *"there are a huge amount of papers declaring what should happen yet it has not"* to showcase how the pro-climate change perspective may fail here. In extension, D26 brought up an example of how *"the Fremantle harbor was developed from a shallow delta with a channel that the first explorers sailed up and found the Swan estuary which had freshwater springs and a sand bank and became Perth."* In this exchange with our intervention, the deniers showed a more open attitude of sharing their own thoughts and their own set of evidence of why they believed in the conspiracy theories instead of fixating on the specific terminologies and elevating the discussion to an emotional level.

4.2.3 Responding depending on original poster's emotion. Using EmoBERT to test our data's emotional level (Figure 4), we found that if deniers (15/57) expressed a neutral and/or curious attitude towards climate change denialism instead of deeply invested in them, they would be more likely to engage in a back-and-forth knowledge exchange with our bot account. And after the interaction with our intervention, 11 out of the 15 deniers (who showed a neutral and/or curious attitude) maintained their neutral and/or curious attitude, demonstrating that our interventions kept them open to conversations and further engagement. Additionally, 4 out of the 15 deniers expressed a more positive attitude after the interaction, showing a possible increase in how open they are to evidence and further conversations. For instance, when our intervention showed the *"current rate of sea level rise"* with linked evidence, D26 who was asking questions about how to think about the *"polar bears which were meant to be in decline and they are not,"* and asked *"how do I think about this problem"* responded to our evidence with a well-defined discussion. Similarly, in an exchange with D2, we found that when the deniers held a more curious attitude, they are more likely to engage in an intellectual conversation. When D2 cited evidence explaining no increase in natural disasters
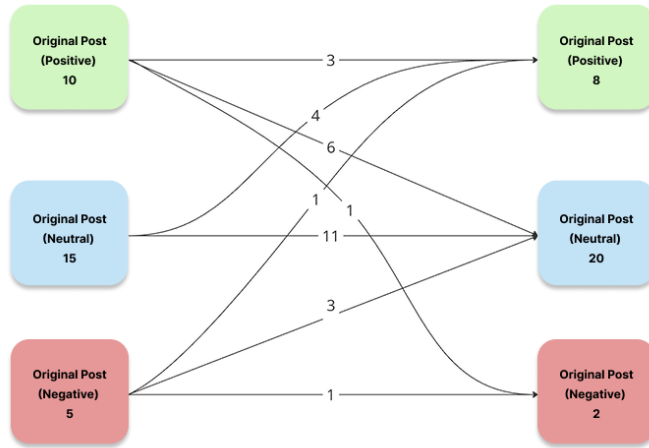
Fig. 4. This image shows the EmoBERT analysis result of the comment (original post) and the response to the corresponding intervention. For the original post and its corresponding responses, we grouped the analysis result into three categories: positive (e.g., optimism, curiosity, etc.), neutral, and negative (e.g., anger, annoyance, etc.). We considered the negative posts as a reflection of the deniers showing a strong emotional investment in the conspiracy theories. In contrast, if the posts expressed a neutral or positive attitude, we considered that the deniers expressed a low emotional investment. Note that this graph only showcases the analysis of climate change deniers' original posts.

causally lead to no climate change, D2 used the phrase *"Can you show me where he is wrong?"* When we provided linked evidence, D2 responded by engaging with the citation, explaining how our evidence failed to address the issue: *"Yearly changes in sea level are below the measurement accuracy of the instruments. This citation is saying ice is melting at the fastest rate and yet no one can point to any place on earth."* Though D2 was still entrenched in climate change denialism, D2 was willing to engage in a discourse with our post.

Based on Figure 4, in most cases, deniers' emotional state improved or remained the same. However, one case came to our attention where the denier's emotion changed from positive to negative after encountering our intervention. Initially, D40 was sharing a detailed explanation about how the rise in population caused the temperature rise instead of climate change. In our intervention, we mentioned that *"There is a consensus among scientists that greenhouse gases play a significant role in climate change."* Then D40 responded with *"There is no consensus, only bought-off scientists."* This reinforced our result discussed in subsubsection 4.3.2 that when intervention contained certain terminologies, deniers may be triggered and respond emotionally.

4.3 RQ2: How do climate change supporters respond to our interventions?

Our analyses did not show any direct correspondence between the original post, our intervention, and the corresponding responses (e.g., whether the intervention contained evidence, whether the original poster was emotional, or whether the intervention contained certain terminologies). But we did observe that climate change supporters generally showed a positive attitude toward our intervention. Our intervention allowed them to more easily discuss their pro-climate change perspectives and offered them the opportunity to provide more evidence. Examples of climate change supporters' responses are shown in Table 2.

4.3.1 Responding with more rationale. Climate change supporters in the community used our intervention as a way to extend the discussions of the pro-climate change perspective. When S12 was arguing for the green movement, our intervention introduced the topic of greenhouse gases' role in the Earth's temperature and climate. S12 then asked the following question: *"Tell me what you know about greenhouse gases."* The response demonstrated S12's willingness to share more thoughts on the given topic. Furthermore, when we introduced pro-climate change evidence, S2 built on our intervention and discussed how the article explained the greenhouse effect. Using phrases such as *"that passage provides a great historical account of,"* S2 shared additional insights about how new evidence such as the article *"reflects the evolving nature of scientific inquiry and the refinement of theories."* Similarly, when we introduced evidence about greenhouse gases and a climate change denier (D52) in the community responded with *"There's no consensus. Only among bought-off scientists."* S5, a pro-climate change Redditor, stepped in and argued in our favor: *"No there's a consensus. Whether the consensus is 97% of climate scientists or 88% that's still a very strong consensus."* In this interaction, our intervention received both climate change deniers' and supporters' attention, which encouraged more discussions and debates in the community.

4.3.2 Responding with more pro-climate change evidence. Our intervention also allowed climate change supporters to offer more evidence in follow-up discussions. When our response highlighted "alarmist messages" in response to a pro-climate change argument around temperature change, S8 responded with *"The situation is alarming. We should all be alarmed"* followed by a list of detailed examples of temperature change, such as *"Canada's wildfire as of today have already burned more land than in any previous wildfire season,"* and *"Texas was under a 45-degree Celsius heat dome for two weeks, killing dozens of people."* In this interaction, the climate change supporters built on our intervention to reinforce their own argument. Similarly, when our intervention corroborated S10's post on sea levels are rising by asking the question of *"Can we discuss the current trend of sea level rise and its implications for climate change? The rate of change today is much faster ... indicating the impact of climate change."* S10 showed a positive attitude and shared more evidence supporting how *"ice levels are connected."* Further, as cited sources are provided, N10 also followed up with descriptions and rationales: *"the ice caps reflect some sunlight in the visible range. As there is less ice, less gets reflected and so total warming increases."* Overall, our engagement allowed the climate change supporters to provide more evidence and reinforce their argument in the climate change debate.

5 Discussion

In this work, we studied how to intervene in conspiracy theory communities and how these interventions may influence the way community members engage in climate change debate. Our work took a novel approach to design and deploy the interventions. By interacting with two groups of users, climate change deniers and climate change supporters, we identified what types of interventions worked well in these communities. We also gained insights about how climate change supporters engage in these conspiracy theory communities, pointing out how future work could study and engage with this population.

5.1 Designing Conspiracy Theory Interventions

5.1.1 Community-centric Interventions. In online communities such as Reddit, the community as a whole shapes the way that people speak about the conspiracy theory and how they interact with one another around it. Many existing platforms such as Twitter/X has focused on engaging with the spread of misinformation from an individualistic perspective, where each account and each post is addressed as misinformation. Furthermore, many existing works studying the design of social

interventions focus on targeting individual users as well [6]. Our work instead took an approach to engage the community as a whole because the social context (where people encounter conspiracy theories and how others engage with them in that context) is an important aspect contributing to intervention design [6, 13, 15, 45, 58]. We demonstrated the possibility and importance of thinking about groups of conspiracy theorists as communities. We tried to design our interventions as closely related to the community as possible by engaging with insider language unique to that community. We also used GPT-4 to generate variations of the interventions so that the community members would not feel like they were engaging with an automated account. Additionally, we attempted to simulate the intervention experience as close to interacting with a real member of that community as possible because of the motivated reasoning phenomenon. Prior works suggest that the motivated reasoning phenomenon may occur if the conspiracy theorists experience an identity threat based on the provided evidence [35, 48, 49, 55, 61]. In that case, conspiracy theory believers are more likely to engage their defense mechanism and choose to continue holding their original belief. Our approach engaging with insider language helped us better align with believers' identity and thus avoid the motivated reasoning effect. Our focus on insider language not only enabled us to meet the community where they were, but also helped us understand the discussion themes in these climate conspiracy communities. Overall, our results indicate the benefits of using community-centric interventions to engage with conspiracy theories. However, since our approach used a manual approach to collect and counter the insider terms, future work could explore how to automatically detect and analyze these insider languages unique to each community. This could make the process of designing interventions easier and more efficient.

5.1.2 Evidence-based Interventions. Providing evidence to support the pro-climate change perspectives was an important component of our intervention. Deniers showed a higher tendency to respond and engage when evidence was provided. However, existing literature pointed out that these communities may be prone to having scientific argumentation [90, 91], engaging in evidence provided by others through certain rationales to support their own claims. Future work could deploy interventions in a wide range of conspiracy theory communities to further understand the generalizability of the approach to use evidence. Additionally, when designing our interventions, we carefully considered how to best include the evidence since prior works suggested that the backfire effect [9, 38, 66, 70, 71] could impact the conspiracy theory believers' consumption of information. As existing works noted [9, 38, 66, 70, 71], the backfire effect occurs when people fail to consume information objectively, which means they are less likely to change their opinions about the original information. Thus, we fine-tuned our interventions to be as respectful and polite as possible when raising opposing opinions and evidence. Future works intervening in these communities should also account for the possible impact of backfire effect when introducing evidence.

5.2 Intervening climate change deniers

In our study, we deployed our interventions in subreddits and observed how users interacted with our posts. After analyzing the collected data, we observed that there are two groups of Redditors in the community: climate change deniers and supporters.

Our results indicated that when interacting with climate change deniers, the most important component of an intervention is linked evidence. When climate change deniers engaged with our interventions, they directly engaged with the provided evidence. In fact, deniers often explicitly ask for evidence if someone is proposing an opposing view. Without the evidence in the intervention, more back-and-forth discussions about what the evidence is and where the evidence is from would ensue. Our results echo prior research in other domains (e.g., health, psychology, education, etc.) [7, 50, 54], showcasing that evidence is a key component in communicating ideas with others,

especially if the goal is to intervene others' perspectives and the evidence is introduced in a respectful way.

When deploying the interventions, another important aspect is to avoid certain terms such as "climate change scientists." Explicitly mentioning the pro-climate change perspectives could lead to a complete shutdown towards any evidence and rationale posted in an intervention. Additionally, we found that how emotionally invested the original poster was, influenced how the poster responded. As explained in subsubsection 4.2.3, when the emotional investment was higher, there was a higher likelihood that the person would respond negatively. We also observed that climate change deniers were willing to share their personal stories to corroborate their point of view. Prior research has shown that personal stories can invoke emotional resonance with the audience [37]. Leveraging existing discussions about how to share a personal story [62] and highlight human values [22] in response to conspiracy theories, future research could explore how to best use an automated system to generate such interventions. Additional research could also investigate how to increase people's exposure to alternative points of view. For instance, an automatic system could be developed to respond to conspiracy theorists even after they started reacting emotionally. This approach could collect more insights about how conspiracy theorists engage in an emotional state, and help explore how to best engage with them in these situations. Moreover, our work indicated that whether the conspiracy theory believers are emotionally invested is an indicator of how they would respond to interventions. Prior works have also suggested that "emotional selection" influences how people consume content [84]. Specifically, conspiracy theory believers are more likely to engage with content that correspond to their emotional state. Similarly, Heath et al. suggested that emotion snowballing could occur when intense emotion is spread from one another [43]. To more closely investigate how emotion influences people's perceptions of the interventions, future work could conduct experiments to compare how people respond across situations of emotional selection, emotion snowballing and so on. Another aspect worth exploring is how to sway the "movable middle." [77] The "movable middle" are people who have not decided about whether they would believe in the conspiracy theories yet. Future work could study this population specifically and design the interventions more tailored to their perspectives.

5.3 Interacting with climate change supporters

Prior studies have identified the existence of a group of users who actively counter misinformation and hate speech online [1, 2, 29, 53, 76]. Our work highlighted a similar group of users who actively engage with conspiracy theories. Specifically, our work revealed how pro-climate change Redditors interacted with members of conspiracy theory communities and how our interventions supported their efforts by providing additional evidence for climate change. Our work underscored that people who do not hold the belief of the community may play an important role in encouraging more open discussions in the community. Future work can more specifically focus on counter-attitudinal populations, or pro-climate science populations, and explore how to best support them in engaging effectively in conversations in conspiracy theory communities. They could explore the specific needs of these users, including how to support their process of collecting information, how to support their process of organizing that information, and how to support their process of interacting with conspiracy theory believers. For instance, based on Twitter/X's Birdwatch project and the Dangerous Speech Project [1, 2], there is an interesting opportunity to explore how to aggregate a central hub of information countering conspiracy theories to help alleviate these users' burdens.

6 Ethical Considerations

Throughout the design and deployment of the study, we tried to minimize harm. Specifically, we were interested in engaging with conspiracy theory community members to explore whether and

how to change their beliefs. To do so, we needed to build trust with the community members. However, this is difficult to do because these communities do not trust scientists at all, which is one reason why they believe in conspiracy theories. Thus, instead of running a full deception study, we chose a middle ground. Building on existing research deploying interventions to misinformation and conspiracy theories [10, 36, 42, 68, 73, 83], when we created the accounts for deployment, we made sure to self-disclose the accounts as bots used for research purposes instead of deceiving the community members. However, since we did not explicitly disclose our accounts as bots when we posted interventions each time (the information that we were researchers was only listed under the account profile), some community members may not have realized that we were conducting research. We believe that future research should also ensure transparency using approaches such as account labeling when deploying interventions. When we collected data from the Reddit communities and used the SAGE method to identify insider language, we made sure all data was anonymized. When we analyzed the collected data (responses to our interventions), we ensured that all users were anonymized. We also made sure that we only collected and analyzed users' posts in the community and did not collect any de-identifiable information from them. Second, when we designed our interventions, we also tried to fine-tune them to be polite and respectful.

7 Limitations

There are some limitations to this work. First, our sample size was relatively small due to several external challenges we faced in the deployment phase, such as the Reddit blackout. Additionally, while lots of our interventions received upvotes and downvotes from community members, some did not receive a response in the text format. Thus, we had a relatively small sample to perform qualitative analysis on. Future large-scale deployment could help resolve these issues and provide more quantitative insights. Additionally, our work only explored how people responded to our intervention. We did not further engage with people who responded with follow-up conversations. Future work should study these follow-up interactions more closely to further understand how to best engage with conspiracy theory believers.

8 Conclusion

In this work, we designed and deployed interventions addressing conspiracy theories in Reddit communities. We designed interventions tailored to the conspiracy communities by addressing the specific insider terms they used. Our work highlights the process and challenges of automatically intervening in conspiracy theory communities. Additionally, throughout the deployment, we encountered a group of users who voluntarily engaged with conspiracy theories, highlighting future opportunities to study and engage with them. Overall, our work provides implications for future works studying and interacting with conspiracy theorists.

9 Acknowledgement

References

[1] [n. d.]. Dangerous Speech Project. https://dangerousspeech.org/
[2] 2021. Introducing Birdwatch, a community-based approach to misinformation. Retrieved Jan 21, 2021 from https://blog. twitter.com/en_us/topics/product/2021/introducing-birdwatch-a-community-based-approach-to-misinformation
[3] N Abas and N Khan. 2014. Carbon conundrum, climate change, CO2 capture and consumptions. *Journal of CO2 Utilization* 8 (2014), 39–48.

[4] Mark S Ackerman. 2000. The intellectual challenge of CSCW: the gap between social requirements and technical feasibility. *Human–Computer Interaction* 15, 2-3 (2000), 179–203.

[5] Olanrewaju Tahir Aduragba, Jialin Yu, Alexandra I Cristea, and Lei Shi. 2021. Detecting fine-grained emotions on social media during major disease outbreaks: health and well-being before and during the COVID-19 pandemic. In *AMIA annual symposium proceedings*, Vol. 2021. American Medical Informatics Association, 187.

[6] Zhila Aghajari, Eric PS Baumer, and Dominic DiFranzo. 2023. Reviewing Interventions to Address Misinformation: The Need to Expand Our Vision Beyond an Individualistic Focus. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–34.

[7] Loai Albarqouni, Tammy Hoffmann, and Paul Glasziou. 2018. Evidence-based practice educational intervention studies: a systematic review of what is taught and how it is measured. *BMC medical education* 18, 1 (2018), 1–8.

[8] Michelle A Amazeen, Arunima Krishna, and Rob Eschmann. 2022. Cutting the bunk: Comparing the solo and aggregate effects of prebunking and debunking COVID-19 vaccine misinformation. *Science Communication* 44, 4 (2022), 387–417.

[9] Craig A Anderson, Mark R Lepper, and Lee Ross. 1980. Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of personality and social psychology* 39, 6 (1980), 1037.

[10] Amanda Baughan, Katherine Alejandra Cross, Elena Khasanova, and Alexis Hiniker. 2022. Shame on Who? Experimentally Reducing Shame During Political Arguments on Twitter. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–18.

[11] Salil D Benegal and Lyle A Scruggs. 2018. Correcting misinformation about climate change: The impact of partisanship in an experimental setting. *Climatic change* 148, 1-2 (2018), 61–80.

[12] Jeff Berardelli. 2020. 10 common myths about climate change — and what science really says. https://www.cbsnews.com/news/climate-change-myths-what-science-really-says/

[13] Alessandro Bessi, Mauro Coletto, George Alexandru Davidescu, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. 2015. Science vs conspiracy: Collective narratives in the age of misinformation. *PloS one* 10, 2 (2015), e0118093.

[14] William R Brown. 1982. Attention and the rhetoric of social intervention. *Quarterly Journal of Speech* 68, 1 (1982), 17–27.

[15] Jovan Byford. 2011. *Conspiracy theories: A critical introduction.* Springer.

[16] Anny Cazenave and Frederique Remy. 2011. Sea level and climate: measurements and causes of changes. *Wiley Interdisciplinary Reviews: Climate Change* 2, 5 (2011), 647–662.

[17] Man-pui Sally Chan, Christopher R Jones, Kathleen Hall Jamieson, and Dolores Albarracín. 2017. Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological science* 28, 11 (2017), 1531–1546.

[18] Hans Christian, Mikhael Pramodana Agus, and Derwin Suhartono. 2016. Single document automatic text summarization using term frequency-inverse document frequency (TF-IDF). *ComTech: Computer, Mathematics and Engineering Applications* 7, 4 (2016), 285–294.

[19] Matteo Cinelli, Gabriele Etta, Michele Avalle, Alessandro Quattrociocchi, Niccolò Di Marco, Carlo Valensise, Alessandro Galeazzi, and Walter Quattrociocchi. 2022. Conspiracy theories and social media platforms. *Current Opinion in Psychology* (2022), 101407.

[20] Yale Climate Change Communication. 2023. Global Warming's Six Americas. https://climatecommunication.yale.edu/about/projects/global-warmings-six-americas/

[21] John Cook. 2020. Deconstructing climate science denial. *Research handbook on communicating climate change* (2020), 62–78.

[22] Adam Corner, Ezra Markowitz, and Nick Pidgeon. 2014. Public engagement with climate change: the role of human values. *Wiley interdisciplinary reviews: climate change* 5, 3 (2014), 411–422.

[23] Matthew RX Dentith. 2019. Conspiracy theories on the basis of the evidence. *Synthese* 196 (2019), 2243–2261.

[24] Matthew RX Dentith and Martin Orr. 2018. Secrecy and conspiracy. *Episteme* 15, 4 (2018), 433–450.

[25] Karen M Douglas and Robbie M Sutton. 2015. Climate change: Why the conspiracy theories are dangerous. *Bulletin of the Atomic Scientists* 71, 2 (2015), 98–106.

[26] Karen M Douglas, Robbie M Sutton, and Aleksandra Cichocka. 2017. The psychology of conspiracy theories. *Current directions in psychological science* 26, 6 (2017), 538–542.

[27] Karen M Douglas, Robbie M Sutton, Daniel Jolley, and Michael J Wood. 2015. The social, political, environmental, and health-related consequences of conspiracy theories. *The psychology of conspiracy* 56 (2015), 183–200.

[28] Karen M Douglas, Joseph E Uscinski, Robbie M Sutton, Aleksandra Cichocka, Turkay Nefes, Chee Siang Ang, and Farzin Deravi. 2019. Understanding conspiracy theories. *Political psychology* 40 (2019), 3–35.

[29] Chiara Patricia Drolsbach and Nicolas Pröllochs. 2023. Diffusion of community fact-checked misinformation on Twitter. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–22.

[30] Riley E Dunlap, Aaron M McCright, et al. 2011. Organized climate change denial. *The Oxford handbook of climate change and society* 1 (2011), 144–160.

[31] Catherine Dwyer. 2007. Task Technology Fit, the social technical gap and social networking sites. *AMCIS 2007 Proceedings* (2007), 374.

[32] Jacob Eisenstein, Amr Ahmed, and Eric P Xing. 2011. Sparse additive generative models of text. In *Proceedings of the 28th international conference on machine learning (ICML-11)*. 1041–1048.

[33] Kristen Engel, Shruti Phadke, and Tanushree Mitra. 2023. Learning from the Ex-Believers: Individuals' Journeys In and Out of Conspiracy Theories Online. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–37.

[34] Robert Etkins and Edward S Epstein. 1982. The rise of global mean sea level as an indication of climate change. *Science* 215, 4530 (1982), 287–289.

[35] Daniel J Flynn, Brendan Nyhan, and Jason Reifler. 2017. The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics. *Political Psychology* 38 (2017), 127–150.

[36] Isabelle Freiling, Nicole M Krause, and Dietram A Scheufele. 2023. Science and ethics of "curing" misinformation. *AMA journal of ethics* 25, 3 (2023), 228–237.

[37] Terrence L Gargiulo. 2006. Power of stories. *The journal for quality and participation* 29, 1 (2006), 4.

[38] R Kelly Garrett and Brian E Weeks. 2013. The promise and peril of real-time corrections to political misperceptions. In *Proceedings of the 2013 conference on Computer supported cooperative work*. 1047–1058.

[39] Fabien Girardin. 2007. *Towards Reducing the Social-Technical Gap in Location-Aware Computing*. Ph. D. Dissertation. Ph. D. Dissertation. Citeseer.

[40] Andreas Goreis, Oswald D Kothgassner, et al. 2020. Social media as vehicle for conspiracy beliefs on COVID-19. *Digital Psychology* 1, 2 (2020), 36–39.

[41] Robin Lambert Graham, Monica G Turner, and Virginia H Dale. 1990. How increasing CO2 and climate change affect forests. *BioScience* 40, 8 (1990), 575–587.

[42] Dominik Hangartner, Gloria Gennaro, Sary Alasiri, Nicholas Bahrich, Alexandra Bornhoft, Joseph Boucher, Buket Buse Demirci, Laurenz Derksen, Aldo Hall, Matthias Jochum, et al. 2021. Empathy-based counterspeech can reduce racist hate speech in a social media field experiment. *Proceedings of the National Academy of Sciences* 118, 50 (2021), e2116310118.

[43] Chip Heath, Chris Bell, and Emily Sternberg. 2001. Emotional selection in memes: the case of urban legends. *Journal of personality and social psychology* 81, 6 (2001), 1028.

[44] Reddit Help. [n. d.]. What is karma? https://support.reddithelp.com/hc/en-us/articles/204511829-What-is-karma-

[45] Richard Hofstadter. 2012. *The paranoid style in American politics*. Vintage.

[46] Eklas Hossain. 2023. *The sun, energy, and climate change*. Springer.

[47] Mark Jary. 1998. Relevance theory and the communication of politeness. *Journal of pragmatics* 30, 1 (1998), 1–19.

[48] Dan M Kahan. 2013. Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision making* 8, 4 (2013), 407–424.

[49] Joseph Kahne and Benjamin Bowyer. 2017. Educating for democracy in a partisan age: Confronting the challenges of motivated reasoning and misinformation. *American educational research journal* 54, 1 (2017), 3–34.

[50] Susan M Kegeles, Gregory Rebchook, Lance Pollack, David Huebner, Scott Tebbetts, John Hamiga, David Sweeney, and Benjamin Zovod. 2012. An intervention to help community-based organizations implement an evidence-based HIV prevention intervention: The mpowerment project technology exchange system. *American Journal of Community Psychology* 49, 1-2 (2012), 182–198.

[51] Taewoon Kim and Piek Vossen. 2021. Emoberta: Speaker-aware emotion recognition in conversation with roberta. *arXiv preprint arXiv:2108.12009* (2021).

[52] Colin Klein, Peter Clutton, and Adam G Dunn. 2019. Pathways to conspiracy: The social and linguistic precursors of involvement in Reddit's conspiracy theory forum. *PloS one* 14, 11 (2019), e0225098.

[53] Alex Zhi-Xiong Koo, Min-Hsin Su, SangWon Lee, So-Yun Ahn, and Hernando Rojas. 2021. What motivates people to correct misinformation? Examining the effects of third-person perceptions and perceived norms. *Journal of Broadcasting & Electronic Media* 65, 1 (2021), 111–134.

[54] Thomas R Kratochwill and Elisa Steele Shernoff. 2003. Evidence-based practice: Promoting evidence-based interventions in school psychology. *School Psychology Quarterly* 18, 4 (2003), 389.

[55] Ziva Kunda. 1990. The case for motivated reasoning. *Psychological bulletin* 108, 3 (1990), 480.

[56] Elżbieta Kużelewska and Mariusz Tomaszuk. 2022. Rise of conspiracy theories in the pandemic times. *International Journal for the Semiotics of Law-Revue internationale de Sémiotique juridique* 35, 6 (2022), 2373–2389.

[57] Stephan Lewandowsky, John Cook, Ullrich Ecker, Dolores Albarracín, Panayiota Kendeou, Eryn J Newman, Gordon Pennycook, Ethan Porter, David G Rand, David N Rapp, et al. 2020. The debunking handbook 2020. (2020).

[58] Stephan Lewandowsky, Ullrich KH Ecker, and John Cook. 2017. Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of applied research in memory and cognition* 6, 4 (2017), 353–369.

[59] Stephan Lewandowsky, Naomi Oreskes, James S Risbey, Ben R Newell, and Michael Smithson. 2015. Seepage: Climate change denial and its effect on the scientific community. *Global Environmental Change* 33 (2015), 1–13.

[60] Richard S Lindzen. 1997. Can increasing carbon dioxide cause climate change? *Proceedings of the National Academy of Sciences* 94, 16 (1997), 8335–8342.

[61] Milton Lodge and Charles S Taber. 2013. *The rationalizing voter*. Cambridge University Press.

[62] Jonathan Lynn. 2018. Communicating the IPCC: challenges and opportunities. *Handbook of Climate Change Communication: Vol. 3: Case Studies in Climate Change Communication* (2018), 131–143.

[63] Lauren Mackenzie and Megan Wallace. 2011. The communication of respect as a significant dimension of cross-cultural communication competence. *Cross-Cultural Communication* 7, 3 (2011), 10–18.

[64] Nobuo Mimura. 2013. Sea-level rise caused by climate change and its implications for society. *Proceedings of the Japan Academy, Series B* 89, 7 (2013), 281–301.

[65] Stephen A Montzka, Edward J Dlugokencky, and James H Butler. 2011. Non-CO2 greenhouse gases and climate change. *Nature* 476, 7358 (2011), 43–50.

[66] Patricia Moravec, Randall Minas, and Alan R Dennis. 2018. Fake news on social media: People believe what they want to believe when it makes no sense at all. *Kelley School of Business research paper* 18-87 (2018).

[67] Melinda Wenner Moyer. 2019. Why we believe conspiracy theories. *Scientific American* 320, 3 (2019), 58–63.

[68] Kevin Munger. 2017. Tweetment effects on the tweeted: Experimentally reducing racist harassment. *Political Behavior* 39 (2017), 629–649.

[69] Kenzo Nera, Youri L Mora, Pit Klein, Antoine Roblain, Pascaline Van Oost, Julie Terache, and Olivier Klein. 2022. Looking for ties with secret agendas during the pandemic: Conspiracy mentality is associated with reduced trust in political, medical, and scientific institutions–but not in medical personnel. *Psychologica Belgica* 62, 1 (2022), 193.

[70] Raymond S Nickerson. 1998. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology* 2, 2 (1998), 175–220.

[71] Brendan Nyhan and Jason Reifler. 2010. When corrections fail: The persistence of political misperceptions. *Political Behavior* 32, 2 (2010), 303–330.

[72] James Painter and Teresa Ashe. 2012. Cross-national comparison of the presence of climate scepticism in the print media in six countries, 2007–10. *Environmental research letters* 7, 4 (2012), 044005.

[73] Gordon Pennycook, Jonathon McPhetres, Yunhao Zhang, Jackson G Lu, and David G Rand. 2020. Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological science* 31, 7 (2020), 770–780.

[74] Shruti Phadke, Mattia Samory, and Tanushree Mitra. 2021. Characterizing social imaginaries and self-disclosures of dissonance in online conspiracy discussion communities. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–35.

[75] Shruti Phadke, Mattia Samory, and Tanushree Mitra. 2022. Pathways through conspiracy: the evolution of conspiracy radicalization through engagement in online conspiracy discussions. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 16. 770–781.

[76] Nicolas Pröllochs and Stefan Feuerriegel. 2023. Mechanisms of true and false rumor sharing in social media: Collective intelligence or herd behavior? *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–38.

[77] Jan-Willem Prooijen. 2018. *The psychology of conspiracy theories*. Routledge.

[78] Tiago Rocha-Silva, Conceição Nogueira, and Liliana Rodrigues. 2023. Passive data collection on Reddit: a practical approach. *Research Ethics* (2023), 17470161231210542.

[79] Jon Roozenbeek, Sander Van Der Linden, Beth Goldberg, Steve Rathje, and Stephan Lewandowsky. 2022. Psychological inoculation improves resilience against misinformation on social media. *Science Advances* 8, 34 (2022), eabo6254.

[80] Marina Ryabova. 2015. Politeness strategy in everyday communication. *Procedia-Social and Behavioral Sciences* 206 (2015), 90–95.

[81] Mattia Samory and Tanushree Mitra. 2018. Conspiracies online: User discussions in a conspiracy community following dramatic events. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 12.

[82] Mattia Samory and Tanushree Mitra. 2018. 'The Government Spies Using Our Webcams' The Language of Conspiracy Theories in Online Discussions. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–24.

[83] Alexandra A Siegel and Vivienne Badaan. 2020. #No2Sectarianism: Experimental approaches to reducing sectarian hate speech online. *American Political Science Review* 114, 3 (2020), 837–855.

[84] Cass R Sunstein and Adrian Vermeule. 2009. Conspiracy theories: Causes and cures. *Journal of political philosophy* 17, 2 (2009), 202–227.

[85] Kim-Pong Tam and Hoi-Wing Chan. 2023. Conspiracy theories and climate change: A systematic review. *Journal of Environmental Psychology* (2023), 102129.

[86] Christoph Treude and Margaret-Anne Storey. 2009. How tagging helps bridge the gap between social and technical aspects in software development. In *2009 IEEE 31st International Conference on Software Engineering*. IEEE, 12–22.

[87] David Tuffley. 2009. Mind the gap. *International Journal of Sociotechnology and Knowledge Development (IJSKD)* 1, 1 (2009), 58–69.

[88] Joseph E Uscinski. 2019. *Conspiracy theories and the people who believe them.* Oxford University Press, USA.

[89] Joseph E Uscinski, Karen Douglas, and Stephan Lewandowsky. 2017. Climate change conspiracy theories. In *Oxford research encyclopedia of climate science.*

[90] Morgan Wofford. 2022. Parasitic knowledge infrastructures: Data reuse by anthropogenic climate change skeptics. *Proceedings of the Association for Information Science and Technology* 59, 1 (2022), 837–839.

[91] Morgan F Wofford and Andrea K Thomer. 2023. Curating for Contrarian Communities: Data Practices of Anthropogenic Climate Change Skeptics. *Proceedings of the Association for Information Science and Technology* 60, 1 (2023), 442–455.

A Deploying Interventions on Reddit

Deploying counternarratives in online communities, especially communities like r/climateskeptics that reject mainstream scientific consensus is challenging. Conspiracy theorists often distrust established institutions, including science, government, and media [77]. They may view mainstream information sources as part of a conspiracy, making it challenging for researchers to introduce scientific evidence into the community. These challenges we encountered might provide important lessons for future research.

A.1 Adapting interventions across subreddits

As discussed in subsubsection 3.5.1, when deploying our intervention that originally designed for *r/climateskeptics* in the community *r/conspiracy*, some interventions misfired and were out of context (Figure 3b). This was because the two communities, though both discussed climate change conspiracy theories, used the same terms in different ways. For instance, as shown in Figure 3b, members of *r/conspiracy* were discussing the use of "chaos theory," but they did not refer to it in the context of climate change. This mismatch between the terminology and the context it was used in led to our intervention (about climate change) being misfired. Thus, future works exploring intervening in conspiracy theory-related communities should be aware of the uniqueness of each community, and adapt their interventions accordingly. This can help ensure the preciseness and effectiveness of any intervention deployed.

A.2 Account bans

The Reddit accounts used for delivering interventions in the form of replies were transparently marked as bot accounts in the account description. Moreover, to assess the general community norms before starting the experiment, we ensured that both Reddit communities, r/climateskeptics, and r/conspiracy saw engagement from other bot accounts that are not banned. Despite this, several of our accounts got banned from r/climateskeptics without explanations. Our attempts to discuss the bans with community moderators were met with no response. Because of the bans, we could not deliver a large number of interventions to test the statistical significance of various intervention configurations on user responses. We believe that the account bans may be a reflection of how the climate change denialist community may reject external attempts to discuss scientific evidence supporting climate change. We appealed several of our account bans and found that working with the moderation mechanism on Reddit is difficult, specifically in r/climateskeptics and r/conspiracy. It is possible that bots may not be deemed as trustworthy messengers of counter-attitudinal information, even when efforts were made to signal the bots' credibility with karma points. Our experience highlights the importance of building trust between scientists and the online communities discussing conspiracy theories.

A.3 Reddit API changes and Reddit Blackout

In April 2023, Reddit announced changes in the API and data collection policies, which threatened the data access and access to moderation tools from third-party apps. Moderators of many subreddits leveraged their subreddits en masse to protest the API changes. Almost 7000 subreddits went private on June 12th notably including large subreddits such as r/funny, r/gaming, and r/science. The blackouts continued intermittently throughout June making for an unstable intervention environment. While r/climateskeptics didn't directly participate in the blackouts, the instability on the rest of the platform may have affected the traffic, participation, and general discussions on r/climateskeptics. We chose to continue the experiment throughout the blackouts because of uncertainty about the blackout timeline. While intervention experiments are best run in a

stable social media environment, future researchers should consider intervention methods that are resilient to the constantly evolving landscape and monetization models of social media platforms.