[Inner endpaper]

HEC MONTRÉAL

The Effectiveness of Warnings on Multimodal Disinformation

par
Caroline Lussier-Daigneault

Yany Grégoire
HEC Montréal
Codirecteur de recherche

Sylvain Sénécal
HEC Montréal
Codirecteur de recherche

Sciences de la gestion
Spécialisation Expérience Utilisateur

Mémoire présenté en vue de l'obtention
du grade de maîtrise ès sciences en gestion
(M. Sc.)

August 2025

HEC MONTRÉAL

Comité d'éthique de la recherche

CERTIFICATE OF ETHICS APPROVAL

This is to confirm that the research project described below has been evaluated in accordance with ethical conduct for research involving human subjects, and that it meets the requirements of our policy on that subject.

Project No.: 2025-6240

Title of research project: The Effect of Multimodality on Content Credibility

Principal investigator: Caroline Lussier-Daigneault

Co-researchers: Marie Louise Radanielina Hita

Date of project approval: December 03, 2024
Effective date of certificate: December 03, 2024
Expiry date of certificate: December 01, 2025

Maurice Lemelin
Président
CER de HEC Montréal

Signé le 2024-12-03 à 14:16

HEC MONTRÉAL

Comité d'éthique de la recherche

ATTESTATION D'APPROBATION ÉTHIQUE COMPLÉTÉE

La présente atteste que le projet de recherche décrit ci-dessous a fait l'objet des approbations en matière d'éthique de la recherche avec des êtres humains nécessaires selon les exigences de HEC Montréal.

La période de validité du certificat d'approbation éthique émis pour ce projet est maintenant terminée. Si vous devez reprendre contact avec les participants ou reprendre une collecte de données pour ce projet, la certification éthique doit être réactivée préalablement. Vous devez alors prendre contact avec le secrétariat du CER de HEC Montréal.

Projet # : 2025-6240 - Multimodal disinformation

Titre du projet de recherche : The Effectiveness of Warnings on Multimodal Disinformation

Chercheur principal : Caroline Lussier-Daigneault

Cochercheurs : Marie Louise Radanielina Hita

Directeur/codirecteurs : Yany Grégoire; Sylvain Sénécal; Marie Louise Radanielina Hita

Date d'approbation initiale du projet : December 03, 2024
Date de fermeture de l'approbation éthique : August 07, 2025

Maurice Lemelin
Président
CER de HEC Montréal

Signé le 2025-08-07 à 16:59

Résumé

Cette recherche étudie l'effet des avertissements sur la crédibilité perçue de la désinformation et sur la volonté d'interagir avec celle-ci. Elle analyse également si cet effet varie selon la modalité médiatique (vidéo, audio ou texte), un aspect encore peu étudié dans la littérature. Une expérience en ligne intersujets (N = 533) a été menée, répartissant aléatoirement les participants dans l'une des trois modalités. La moitié a reçu un avertissement avant d'être exposée à de la désinformation sur les changements climatiques, tandis que l'autre moitié a vu le même contenu sans avertissement. Après l'exposition au contenu, les participants ont rempli un questionnaire mesurant la crédibilité perçue, la volonté d'interagir avec le contenu, ainsi que des variables complémentaires telles que l'orientation politique, les attitudes environnementales et la littératie numérique.

Les résultats indiquent que les avertissements réduisent significativement la crédibilité perçue et l'intention d'interaction, la crédibilité jouant un rôle médiateur dans cette relation. Contrairement aux attentes, la modalité n'a pas modéré cet effet, les avertissements étant tout aussi efficaces en format vidéo, audio et texte. Ces résultats démontrent que même des avertissements textuels génériques peuvent réduire efficacement la crédibilité perçue et l'intention d'interagir avec de la désinformation. Ils remettent également en question l'hypothèse que la modalité influence fortement les jugements de crédibilité en présence d'avertissements, suggérant qu'un scepticisme généralisé envers le contenu en ligne pourrait surpasser les heuristiques liées au format médiatique.

Mots clés : Désinformation, modalité médiatique, crédibilité perçue, engagement sur les réseaux sociaux, avertissement, inoculation

Méthodes de recherche : Intra-sujets, sondage expérimental quantitatif, mesures autodéclarées

Abstract

This research investigates whether warning labels reduce the perceived credibility of disinformation and the willingness to engage with it across different content modalities (video, audio, and text). A between-subject online experiment (N = 533) was conducted, where participants were randomly assigned to one of three modality conditions. Half of the participants received a generic warning before viewing a piece of climate change disinformation, while the other half viewed the same content without a warning. After exposure, participants completed a self-reported questionnaire measuring perceived credibility and willingness to engage with the content. Additional measures included political affiliation, environmental attitudes, and science media literacy.

Results indicate that the presence of a warning significantly reduced both credibility and engagement, with credibility mediating the relationship between warnings and engagement. Contrary to expectations, modality did not moderate this effect as warnings were equally effective across video, audio, and text formats. These findings contribute to disinformation research by demonstrating that even generic, text-based warnings can effectively reduce the credibility and intent to interact with misleading content. They also challenge the assumption that modality significantly shapes credibility judgments in the presence of warnings, suggesting that generalized skepticism towards social media content may override heuristic cues.

Keywords : Disinformation, modality, perceived credibility, willingness to engage, social media, warnings, inoculation

Research methods : Between-subject, experimental quantitative survey, self-reported measures

Table of contents

Résumé....2
Abstract....3
Table of contents....4
List of tables and figures....7
List of Figures....7
List of Tables....7
List of abbreviations and acronyms....9
Preface....10
Generative AI Usage....10
Acknowledgements....11
Chapter 1
Introduction....1
1.1 The Problem of Disinformation....1
1.2 The Role of Social Media....3
1.3 Purpose and Research Questions....6
1.4 Key Definitions....7
1.5 Contributions....8
1.7 Structure of the Thesis....9
Chapter 2
Literature Review....11
2.1 The Victims of Disinformation....11
2.1.1 Who Believes in Disinformation....11
2.2. Why People Believe in Disinformation....14
2.2.1 The Heuristic-Systematic Model....14
2.3 The Dissemination of Disinformation....18
2.3.1 Social Media Engagement....18
2.4 Warning Interventions and Disinformation....21
2.4.1 Types of Warnings....23
2.5 Engagement and Credibility....24
2.6 Disinformation and Multimodality....26
2.6.1 The Moderating Role of Multimodality....26
2.7 Conceptual Model....28
Chapter 3
Methodology....30
3.1 Pretest....30
3.1.1 Experimental Design....30

3.1.2 Sample....31
3.1.3 Procedure....31
3.1.4 Measures....32
3.1.5 Results....33
3.2 Main Study....35
3.2.1 Experimental Design....35
3.2.2 Context: Climate-Change Disinformation....37
3.2.3 Sample....39
3.2.4 Procedure....41
3.2.5 Measures....42
Chapter 4
Results....44
4.1 Measurements Reliability and Control Variables....44
4.2 Descriptive Statistics....44
4.2.1 Credibility....44
4.2.2 Engagement....45
4.3 The Effect of Warnings on Engagement....47
4.4 The Indirect Effect of Warning on Engagement Through Credibility....48
4.5 Moderation Effect of Modality....51
4.6 Post Hoc Analysis....53
4.6.1 Engagement Across Modalities....53
4.6.2 Item-Level Effects on Engagement....55
4.6.3 Credibility Across Modalities....55
4.6.4 Behavioral Engagement....57
Chapter 5
Discussion....59
5.1 Theoretical Implications....60
5.1.1 Effectiveness of Warnings and HSM....60
5.1.2 Absence of Modality Effects....62
5.1.3 Political Affiliation and Disinformation....65
5.2 Practical Implications....66
5.2.1 General Effectiveness of Warnings and Content Moderation Strategies....67
5.2.2 Modality-Agnostic Nature of Warnings....68
5.2.3 Warning Design....68
5.2.4 Engagement Motivation and Behavioral Differences....69
5.3 Limitations and Future Research....70
Chapter 6
Conclusion....72
6.1 Theoretical Contributions....72

6.2 Practical Contributions....73
6.3 Limitations and Future Research....74
Bibliography....77
Appendices....90
A. Qualtrics Questionnaire (Pretest)....90
1. Introduction and Instructions....90
2. Stimuli....90
3. Comprehension Questions....90
4. Credibility Scale....91
5. Behavioral Engagement Variable....92
6. Engagement Scale....92
7. Controls....93
8. Content Quality....93
9. Demographics....94
10. Environmental Attitude Scale....96
B. Questionnaire (Main Study)....97
1. Introduction....97
2. Instructions....97
3. Comprehension Check....97
5. Behavioral Engagement Variable....98
6. Engagement Scale....98
7. Control....99
8. Demographics....100
9. Environmental Attitude Scale....101
10. Science Media Literacy Scale....101
C. Debrief (Pretest)....103
D. Debrief (Main Study)....105
E. Credibility Scale (Appelman & Sundar, 2016)....106
F. Engagement Scale....107
Pretest....107
Main study....107
G. Environmental Attitude Scale (Milfont & Duckitt, 2010)....108
H. Science Media Literacy (Austin et al., 2023)....116

List of tables and figures

List of Figures

Figure 1. Conceptual Model 29

Figure 2. Video Modality 36

Figure 3. Text Modality 36

Figure 4. Audio Modality 36

Figure 5. Warning and Instructions 37

Figure 6. Correlation Between Engagement and Perceived Credibility 47

Figure 7. Estimated Marginal Means of Engagement Across Warning Conditions, by Modality 49

Figure 8. Validated Moderated Mediation Model 52

Figure 9. Estimated Marginal Means of Engagement Across Warning Conditions, by Modality 54

List of Tables

Table 1. Descriptive Statistics of Pretest Results 34

Table 2. Sample Size by Experimental Condition 40

Table 3. Descriptive Statistics for Perceived Credibility by Modality and Warning Condition 45

Table 4. Descriptive Statistics for Willingness to Engage by Modality and Warning Condition 46

Table 5. ANCOVA Results for the Direct Effect of Warning on Engagement 48

Table 6. ANCOVA Results for the Direct Effect of Warning on Credibility 49

Table 7. Summary of Regression Coefficients for Engagement (Model 7) 50

Table 8. Indirect Effects for Each Modality (Model 7) 51

Table 9. Adjusted Means and Standard Errors for Credibility Scores Across Warning Condition, by Modality Pairs 56

Table 10. Logistic Regression of the Likelihood of Wanting More Information 58

List of abbreviations and acronyms

AI: Artificial Intelligence

CRT: Cognitive Reflection Test

HSM: Heuristic-Systematic Model

LC4MP: Limited Capacity Model of Mediated Message Processing

MRT: Media Richness Theory

MAIN: Modality-Agency-Interactivity-Navigability

U.S.: United States

UX: User Experience

Preface

This thesis was completed as part of the Master's program in User Experience (UX) at HEC Montreal. The research project was approved by the Ethics Board of HEC Montreal under project number 2025-6240. All phases of the study underwent thorough ethical review prior to data collection to ensure compliance with the ethical standards outlined in the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2).

All participants provided informed consent electronically and were informed of their right to withdraw from the study at any point without consequence. Given the nature of this thesis, participants were exposed to disinformation content as part of the study stimuli. To mitigate any potential harm, all participants received a thorough debriefing at the end of both the pretest and the main study (see **Appendix C and D**), clarifying the purpose of the research and identifying the disinformation content.

The author declares no conflict of interest.

Generative AI Usage

The author declares that generative AI was used solely for proofreading, improving concision, and suggesting synonyms or alternative wording only after each section had been fully drafted and again upon completion of the full document. In all cases, the author reviewed and approved all suggested changes before incorporating them.

Acknowledgements

First and foremost, I would like to thank Yany Grégoire, Sylvain Sénécal, and Marie Louise Radanielina Hita, my directorial team, for their support, guidance, and generosity in sharing their knowledge throughout the past year. I am also grateful to them for the opportunity to work on such a timely and impactful topic.

Secondly, I want to thank my husband for his constant support, love, and patience during this process, especially while I was away, either physically, mentally, or both. The past few years have been filled with uncertainty, and you have been my rock throughout. I also want to thank my family for welcoming me back into their home with generosity and for being so understanding of my often chaotic schedule.

I am of course deeply grateful to my classmates, who welcomed me so warmly despite our age difference and helped make the first year of our Master's fly by. Michelle, Jia, Maivel, and so many others, thank you for reminding me that UX research is truly a team sport. I miss you and I am so proud of you all!

Beyond the academic contributions, the topic of this thesis matters to me personally. I have been living in the United States for almost a decade and have witnessed firsthand the growing impact of disinformation campaigns here. The rise of anti-intellectualism, which has allowed blatant falsehoods to spread with impunity even at the highest levels of government, is both disheartening and demoralizing. Yet disinformation is nothing new: as early as 1710, Jonathan Swift observed, "Falsehood flies, and the Truth comes limping after it." As researchers, academics, and individuals, we cannot afford to give up the fight, even when the odds are stacked against us. It is our responsibility to keep trying to make this world less misinformed and to challenge disinformation wherever we encounter it.

Chapter 1
Introduction

1.1 The Problem of Disinformation

Following major events such as the latest U.S. elections, the COVID-19 pandemic, and the war in Ukraine, disinformation has emerged as a significant societal challenge and a potential threat to socio-political stability (Pennycook & Rand, 2020; World Economic Forum, 2024). Indeed, besides feeding people misleading, confusing, or false information, disinformation makes the truth harder to believe as everything could be misconstrued as being false or tampered with. This has been leading to a gradual decrease in trust in public institutions, the media, and authority figures (Di Domenico et al., 2021).

This erosion of trust became overly evident during the 2020 COVID-19 pandemic, which saw a growing number of people pushing for alternative, at times dangerous, solutions to what was proposed by public health services, and questioning the gravity of the situation (Pennycook, McPhetres, et al., 2020). The pandemic quickly became politicized, especially in the U.S., and conspiracy theories went as far as convincing a significant number of individuals that the crisis was a government scheme manufactured to reduce the population, amongst other conspiracies (Imhoff & Lamberty, 2020). The spread of conflicting, false or misguided information that constantly evolves has been called an “infodemic” (Mende et al., 2024). The COVID-19 infodemic directly contributed to increased mortality, as disinformation led people to reject vaccines and undermine protective measures, illustrating its real-life consequences (Imhoff & Lamberty, 2020; Islam et al., 2021).

While the COVID-19 pandemic brought new urgency to the issue, disinformation has long been used to manipulate public perception in other domains such as climate change, which will be the focus of the present thesis. Since it first became a topic of public discussion, climate change has been a major target of misleading and false information. This disinformation has been fabricated by skeptics, and commonly bankrolled by polluting corporations wanting to frame the discourse

in a light that is favorable to their industry (Pierre & Neuman, 2021). One of these enduring, well-known tactics was to place the blame on individuals to lessen corporate responsibility, for instance, through the popularization of the concept of carbon footprint. Today, most of the disinformation about climate change is meant to create uncertainty about its anthropogenic origin, its severity, and the effectiveness of proposed solutions (Sethi, 2024). As Diaz Ruiz & Nilsson (2023) explain about disinformation in general, "[t]he strategy involves moving the goalposts regarding what constitutes knowledge." (p.30) The burden of proof is then pushed onto climate scientists, who are forced to spend time attempting to debunk the perpetual questionings they receive from antagonistic parties (Lewandowsky & van der Linden, 2021).

Overall, it appears that high-profile topics that cause a high emotional response are more likely to lead people to knowingly or unknowingly believe false content (Di Domenico et al., 2021; Martel et al., 2020). These topics are often politicized, especially in the U.S., with individuals' allegiance to certain ideologies influencing what they are willing to accept or share (Pennycook & Rand, 2020). Because such topics have tangible impacts on societal behavior, policy decisions, and public safety, they are frequently weaponized to sow discord and chaos, inadvertently or purposely benefiting groups opposed to the targeted individuals or communities.

Scholars have termed the recent tendency for emotions and personal beliefs to outweigh objective facts in individual and public discourses as "post-truth" (Di Domenico et al., 2021), a paradigm that exacerbates biases, thereby making people even more susceptible to disinformation. This phenomenon complicates efforts to prevent or debunk falsehoods by normalizing the discrediting of factual evidence and equating it to personal opinions or feelings. In other words, presenting proof is no longer sufficient; one must communicate in ways that emotionally resonate with the audience and protect their sense of identity in the way the correction is delivered.

1.2 The Role of Social Media

The threat of disinformation extends beyond the erosion of trust in public institutions and the disregard of public health measures during a pandemic; it also poses a non-negligible risk of inciting violence. For instance, in 2018, false rumors shared on WhatsApp about the identity of child kidnappers in India led to multiple mob lynching of individuals who were wrongly accused (Lewandowsky & van der Linden, 2021). Similarly, in 2017, a humanitarian disaster unfolded in Myanmar after disinformation targeting the Rohingya ethnic group was amplified by Facebook's recommendation algorithm. This incited state-condoned, widespread violence against the Rohingya, forcing over 700,000 of them to flee their homes (Amnesty International, 2022). These examples showcase not only the possible effects of disinformation, but the role of social media in its propagation.

According to a recent survey by the Pew Research Center (2024), more than half of American adults use social media to get their news at least occasionally. Among Americans aged 18 to 29, half reported trusting the news they see on social media sites at least sometimes (Gottfried, 2022). In recent years, the most popular platforms for news consumption in the U.S. have been Facebook (33%), YouTube (32%), Instagram (20%) and TikTok (17%), with TikTok showing the largest increase in popularity since 2020 (from 3%) (Pew Research Center, 2024). Unsurprisingly, most people who consume news on social media are under 50 years old (Pew Research Center, 2024).

It is likely that most people who use social media for news consumption, and even people who use social media platforms for its primary intended purpose, have been exposed to disinformation. In Canada, for example, three-quarters of respondents reported encountering suspicious content online in the past 12 months (Statistics Canada, 2024). Yet, only 44% of people in 27 surveyed countries, including only 26% in the U.S. and 37% in Canada, are confident that others can accurately identify misinformation, although a majority believe they can do so themselves (Ipsos, 2023). This confidence gap is noteworthy, as a recent study by Angelucci and Prat (2024) found that only 47% of participants were able to confidently identify

true stories versus false ones, 50% were uncertain, and 3% were confidently incorrect. These findings, however, do not take into account the amplifying and legitimizing effect of social media, which underscores the importance of studying both topics together.

Although disinformation is not a new phenomenon (Aïmeur et al., 2023), social media platforms have facilitated its spread through affordances that enable users to freely upload and share content with their network and with the public (Sundar et al., 2021). Sundar (2008) defines affordances as "capabilities that can shape the nature of content in a given medium." (p.75). In other words, web-based affordances convey the interactive functionalities of a digital platform, primarily through conventional visual cues such as buttons or text-fields. Therefore, the sharing button on Facebook, which allows users to share content created and posted by others within their own network, is one such affordance that has contributed to the widespread dissemination of disinformation.

The ease of posting content on social media, driven by its affordances, eliminates many of the barriers associated with traditional media. As people shift away to online media to consume information, they must act as the sole judges of its credibility. This task is made more difficult by the overwhelming abundance of information online, as well as a proliferation of self-proclaimed experts who can share opinions without verified credentials (Metzger & Flanagin, 2013).

Besides affordances, the myriads of opaque algorithms is another characteristic of social media that impacts the way disinformation is disseminated. Indeed, one of social media's biggest features is its recommender system, which picks what content social media users will passively be exposed to on their "feed" (Narayanan, 2023). The algorithms that power these recommender systems on platforms like Facebook or TikTok are hard to fully understand due to their proprietary nature. However, it is known that they prioritize engagement, often amplifying polarizing and emotionally charged content that generates interaction (Milli et al., 2021). This dynamic inadvertently promotes the spread of disinformation, fostering echo chambers where users are exposed primarily to content aligning with their preexisting beliefs (Di Domenico et al., 2022; Diaz Ruiz & Nilsson, 2023).

Social media, like most digital platforms, also enables users to create and publish content in various formats, such as text, images, and videos. These different formats are referred to as modalities. Modalities are defined as modes of information processing corresponding to the five senses (Fisher et al., 2019). From a technological or communication perspective, a modality typically describes the method through which information is presented and how users interact with it (Sundar, 2008). For example, audiovisual content, like videos, is considered multimodal as it integrates both visual and auditory modalities into a single piece of content.

Modalities are particularly significant in disinformation research because it can take many forms, from text-based articles to sophisticated videos. Most existing work focuses on text-based misleading content, with or without images, likely due to the prevalence of articles in disinformation campaigns (Ecker et al., 2022). However, with the rise of deepfake technology, with which you can create hyper realistic fabricated videos, there has been growing interest in studying how such content influences the reception of disinformation. For example, Shin & Lee (2022) demonstrated that deepfakes are as likely to be believed and shared as real videos, when it confirms their pre-existing beliefs. Conversely, Vaccari & Chadwick (2020) found that although deepfakes are not systematically deceptive, they contribute to a sense of uncertainty in all news found on social media. Similarly, Hameleers et al. (2022) concluded that while deepfakes do not lead to stronger credibility evaluations than textual disinformation, they risk affecting trust in online news more generally.

Despite this interest in deepfake videos, few studies systematically compare the credibility of disinformation across different modalities, especially across text, audio, and videos (Ecker et al., 2022). This leaves unanswered questions on the effects of information mode of presentation on the potential for disinformation to spread and be believed. Addressing this gap is essential for developing effective solutions that account for the multimodal nature of modern disinformation, particularly on social media platforms where users encounter diverse formats daily.

Given the difficulty of combating disinformation, particularly in a post-truth society, social media platforms have attempted various strategies to address the problem. However, many question whether private companies that thrive on engagement will prioritize user well-being over shareholder interests without any government oversight (Diaz Ruiz, 2023). Furthermore, these platforms are perpetually playing catch-up with disinformation, as it spreads faster than it is removed, and ends up walking a fine line between censorship and moderation. While existing research has explored interventions to reduce the effect of disinformation (see Aïmeur et al., 2023; Ecker et al., 2022; Martel & Rand, 2023; Mende et al., 2024), there is limited understanding of how these strategies perform in the context of multimodal content. To bridge this gap, it is crucial to investigate the effectiveness of interventions across different content formats, with the aim of developing guidelines that can inform policymakers and promote the adoption of best practices among private companies.

1.3 Purpose and Research Questions

Given the significant societal impact of disinformation, the role of social media in amplifying its spread, and the lack of research on how modality influences the effectiveness of prevention strategies, this thesis investigates the impact of warnings on different content formats. Specifically, it examines whether warnings can reduce both the credibility of and engagement with climate change disinformation across text, audio, and video modalities. This thesis uses both attitudinal (perceived credibility and engagement intent), and behavioral (interest in receiving more information about the content) measures to evaluate user responses.

Accordingly, this thesis addresses the following research questions:

- To what extent do warnings influence disinformation credibility and engagement intent?
- To what extent does modality influence the relationship between warnings and credibility of disinformation?

To answer these questions, an online experiment was conducted in which the presence of a warning and the content modality were systematically manipulated, providing insights into the effectiveness of warnings across formats.

1.4 Key Definitions

To better situate this thesis in the literature, it is important to clarify the meaning of its key constructs and how they are operationalized. Social media *engagement*, for instance, is related to other forms of engagement such as customer engagement, yet here it specifically refers to behaviors that reflect interaction with content, often as a way to indicate the impact the content had on the individual (Syrdal & Briggs, 2018). In this thesis, engagement is operationalized as the *willingness to engage* (also referred to as *engagement intent*), measured through self-reported intentions rather than observed behaviors.

Credibility, in this context, refers to message credibility, that is, an evaluation of the truthfulness of the content itself rather than its source (Appelman & Sundar, 2016). As it is measured through self-report as well, it is treated as *perceived credibility*, which does not align with an objective assessment of accuracy.

Warnings, as used in this thesis, are visual disclosures of information that can take the form of symbols, brief messages, or a combination of both. Their purpose is to alert viewers about a particular aspect of the content, prompting them to evaluate it with that consideration in mind.

As previously mentioned, *modality* pertains to the mode of information processing and sensory channel through which content is presented. On social media, common modalities include visual, auditory, and textual content, the latter being technically visual but engaging distinct cognitive processes (Lang, 2000). Whereas social media platforms often support multimodal content, combining two or more modalities, certain other media, such as radio, are unimodal (i.e., in this case, relying only on the auditory channel).

Finally, it is also important to clarify the type of information that this study will focus on and distinguish the common terms found in the literature. *Misinformation* usually refers to information that is inaccurate, misleading, or incomplete (Diaz Ruiz & Nilsson, 2023). *Disinformation*, by contrast, is a form of misinformation created with the deliberate intent to deceive people and intentionally distributed to cause harm (Mende et al., 2024). Contrary to popular belief, disinformation is not always completely false; it can include elements of truth to enhance its credibility (Diaz Ruiz & Nilsson, 2023). This is because the goal of disinformation is not necessarily to convince people of falsehoods, but to plant seeds of doubt in their minds.

Disinformation includes *fake news*, a term that gained popularity during 2016 U.S. election (Jones-Jang et al., 2021), which refers to a form of disinformation presented as news stories, and *propaganda*, which is deliberately created by or for political entities to harm the interest of others (Aïmeur et al., 2023). Thus, when a political entity purposely fabricates and disseminates misleading information to sow distrust in a target, such as foreign nations fabricating false stories to influence public opinion in favor of a chosen politician, it constitutes both propaganda and, more generally, disinformation. This thesis focuses specifically on disinformation, given its intentional nature and the significant threat it poses to society. It is also particularly challenging to combat, as malicious actors continuously seek to circumvent safeguards.

1.5 Contributions

This thesis makes several contributions to the theory, practice and methodology in the field of disinformation research. First, it builds on the Heuristic-Systematic Model (HSM) by applying it in the context of disinformation and demonstrating that warnings may trigger accuracy motivation, which encourages individuals to engage in a more analytical systematic processing rather than the more intuitive heuristic processing. It also extends the HSM past attitude change by showing that credibility judgments mediate behavioral intentions, such as the willingness to engage with disinformation content.

Second, this research challenges key assumptions from the MAIN model and Media Richness Theory (MRT), which suggest that sensory-rich media (e.g., video) are inherently more credible and, as a result, may be less responsive to disinformation mitigation solutions. In this study, content modality neither significantly affected credibility nor impacted the effectiveness of warnings on credibility, suggesting that modality-based heuristic cues were minimal and that generic warnings were equally effective across formats in reducing credibility. In contrast, the effect of warnings on engagement did somewhat vary by modality. Warnings had the strongest impact in the audio condition, a smaller impact in video, and no impact in text.

Methodologically, this study contributes to the growing literature on multimodal disinformation by comparing three content modalities (text, audio, and video) within a single experiment embedded in a survey. It also tests warnings formatted directly into the survey, making this approach easy to replicate in future experimental studies.

From a practical standpoint, the results provide evidence that simple, scalable interventions like text-based generic warning labels can help reduce the credibility and engagement with disinformation across formats. These findings are particularly relevant for social media platforms and other digital services where users interact with content, especially considering the widespread and growing influence of disinformation online.

1.7 Structure of the Thesis

This thesis is organized as follows: the next chapter, Chapter 2, reviews the relevant literature on disinformation, and introduces the theoretical foundations that informed the development of the research model and research questions presented in the introduction. It also outlines the hypotheses that serve to structure and guide the research. Chapter 3 covers the research methodology and defines the dependent variables that will be used to measure the outcome of the research model. Chapter 4 presents the results, while Chapter 5 discusses their interpretation in the context of the theoretical framework introduced earlier. Finally, Chapter 6 concludes this

thesis by summarizing the key findings, addressing the limitations, and highlighting the practical implications of the research.

Chapter 2
Literature Review

This chapter reviews the current body of knowledge relevant to the study of disinformation, and outlines the research model and hypotheses, as well as the theory that underpins both. It begins with a discussion of recent studies on disinformation, then examines research on warning interventions and content multimodality in this context.

2.1 The Victims of Disinformation

To understand disinformation better, it is crucial to start by examining the contexts that lead people to believe in disinformation, the type of characteristics present in disinformation believers and spreaders, and the mechanism of its dissemination. Although this thesis focuses on disinformation, it will also review the literature on different types of misinformation, as the constructs are often used interchangeably in the literature and are all generally incorporated into relevant theoretical foundations.

2.1.1 Who Believes in Disinformation

It is difficult to draw a precise portrait of who believes and spreads misinformation. However, research generally agrees that individuals most susceptible to conspiracy thinking or disinformation often feel alienated from society in some way and are experiencing insecurity and uncertainty in their lives (Booth et al., 2024). Most of the time, the disinformation they believe in is at least partially aligned with their existing beliefs. Booth et al. (2024) also note that individuals who become radicalized by disinformation to the point of joining extremist groups are often motivated by a desire to belong. Ironically, as they are funneled into more radicalized circles and discourses, they tend to withdraw from their previous networks and increasingly rely on their newfound community. At that point, these individuals will rationalize the network of information adhered to by their community as a way to demonstrate loyalty and reaffirm their membership (Kahan, 2013).

To illustrate this point, Diaz Ruiz & Nilsson (2023) investigated how disinformation and conspiracies are shared through platforms and people, taking the flat earth movement as a case study. The authors reported that identity reinforcing, through the consumption of disinformation, is a potent vehicle of dissemination. They argue that victims of disinformation are not always passively being duped but can also take an active role by spreading disinformation because it fits with the identity they are constructing and projecting.

The authors found that flat-earthers do not form a cohesive group but instead consist of multiple subgroups with different reasons for believing in the conspiracy. What brings these groups together is a shared sense of identity brought forward by their belief in the conspiracy, which creates an in-group, the believers, and an out-group, the non-believers, and aligns with the radicalization pathway defined by Booth et al. (2024). In other words, the desire to belong to something meaningful will drive individuals to accept radical beliefs if it allows them to claim membership to the community who promotes these beliefs.

Other commonly examined individual characteristics are rationality and emotionality. In their research, Martel et al. (2020) studied the role of emotional processing on the belief in fake news. They found that heightened emotional responses increased the perceived accuracy of fake news and reduced participants' ability to distinguish between real and fake headlines. The specific type or valence of emotion did not make a difference. Since disinformation is often designed to trigger strong emotional reactions (Mende et al., 2024), it can create a feedback loop where the more emotionally reactive the audience, the more emotionally charged the content is designed. Therefore, individuals going through strong emotions, or those who are generally more emotionally reactive, may be more vulnerable to disinformation.

Additionally, research shows a negative correlation between performances on the Cognitive Reflection Test (CRT) and susceptibility to fake news (Pennycook & Rand, 2019). The CRT measures analytical thinking through a series of questions that appear straightforward at first glance but require more thorough analysis to properly answer. Though Pennycook & Rand (2019) only found weak evidence of the correlation between analytic thinking and believing true

news stories, in a later study, they found that people who score higher on the CRT were better at identifying true versus false headlines (Bago et al., 2020). Thus, individuals with a greater predisposition for analytical thinking may be better equipped to resist disinformation.

Similarly, another study by Pennycook & Rand (2020) found that individuals more receptive to pseudo-profound motivational sentences were more likely to see fake news as accurate. This receptivity, along with a tendency to overclaim knowledge or be overconfident in general knowledge familiarity, was negatively correlated with CRT results. The authors suggest that both pseudo-profound statements and fake news demonstrate indifference for the truth, which may explain the relationship between both. Participants who were more receptive to such statements were also more likely to say they would share fake news on social media. Interestingly, the authors found only a modest correlation between the perceived accuracy of real news and the likelihood of sharing it online. They suggest that sharing on social media has more to do with social membership or reputation than the desire to be accurate.

Finally, political orientation has been shown to influence the effectiveness of disinformation labels. Mende et al. (2024), in a review of the literature on disinformation, noted that Democrats are less likely to share dubious content on Twitter, than Republicans. A systematic review of the climate change conspiracy literature by Tam & Chan (2023) also reported that individuals who deny climate change tend to be older conservative men who are more religious, educated and wealthy than climate skeptics. They tend to have lower trust in the media and public authorities as well. The paper further notes that belief in climate change being a hoax is often associated with a broader conspiracy mindset.

Additionally, according to Lewandowsky et al. (2017), backfire effects, a cognitive bias where correcting people pushes them to strengthen their initial belief, were most often observed when correcting information that challenged Republican worldviews. For Democrats, while backfire effects were less common, corrections were still less effective with partisan information. On the other hand, Angelucci & Prat (2024) found that socioeconomic inequality was more important than partisanship to explain participants' ability to identify true and false stories. Taken together,

these findings suggest that while political orientation can indicate who is more susceptible to disinformation, it does not work in isolation.

Overall, vulnerability to disinformation appears to be shaped by a combination of factors such as political alignment, cognitive abilities, emotionality, socioeconomic status, and, importantly, feelings of social isolation. This underscores the need for a multidimensional approach to understanding and addressing disinformation, one that integrates psychological, cognitive, and structural factors of susceptibility.

2.2. Why People Believe in Disinformation

There are many explanations for why people believe in disinformation. One influential framework, the Dual Process theories, is frequently used to explain the cognitive mechanism behind the believability of disinformation. This set of theories suggests that individuals process information through one of two modes: a fast, intuitive mode that tends to be unconscious, or a more effortful, analytical mode. The exact distinctions and underlying mechanisms vary depending on the specific version of the theory being referenced, but the core principles are mostly the same. This thesis will focus on the Heuristic-Systematic Model (HSM) developed by Chaiken (1980).

2.2.1 The Heuristic-Systematic Model

The Heuristic-Systematic Model posits that there are two pathways for processing information: the systematic route, which involves comprehensively analyzing the information and, as such, requires more cognitive effort, and the heuristic route, which relies instead on processing only a subset of the information through cognitive shortcuts (Chaiken, 1980; Chaiken & Ledgerwood, 2012; Todorov et al., 2002). Reliance on cognitive shortcuts, such as judging the credibility of information based on its familiarity or the authority of its source, can explain why knowledge is sometimes overlooked. In other words, if specific stimuli cue the brain to use shortcuts instead of a more effortful analysis, it can skew perceptions of information credibility.

The choice of which mode of information processing to use depends on both internal and external factors (Todorov et al., 2002). The underlying idea is that people are "cognitive misers", who seek to conserve cognitive resources by processing information as quickly and efficiently as possible. This means people will engage in more effortful processing only if enough cognitive resources are available to them at that moment, and if motivation is sufficient to invest more effort into the task at hand. Therefore, when external factors, such as information overload or time constraints, are hindering the availability of cognitive resources and motivation, individuals are more likely to rely on heuristic processing. In fact, heuristic processing tends to be the preferred processing mode in everyday decision-making (Sundar et al., 2021).

It is to be noted that heuristic processing is constrained by the availability of heuristics, or judgment rules, as Sundar (2008) described them. These rules develop over time through experience and exposure to relevant cues. For example, an individual may learn in school that information from a source with a title such as 'Doctor' or 'PhD' is more reliable, thus memorizing a source-expertise shortcut. Both Todorov et al. (2002) and Sundar (2008) emphasize that heuristics are not solely used during heuristics processing as they can also be useful analytical tools during systematic processing when applied willfully.

One of the possible reasons why people may resort to heuristic processing when coming across disinformation on social media is that the abundance of content and the speed at which it changes is overwhelming. This information overload might effectively push people towards a more superficial analysis of the content as the brain does not have the capacity to encode all of the information (Appelman & Sundar, 2016; Lang, 2006). This is in addition to affordances such as infinite scrolling, which has become the norm on most social media platforms and tends to push people to consume more content (American Psychological Association, 2024).

The HSM proposes two assumptions regarding the motivation behind the selection of a processing mode. One of these is explained by the *sufficiency principle*. Per Todorov et al. (2002): "The sufficiency principle conceptualizes motivation to engage in information processing as a function of the discrepancy between the person's actual confidence and the

person's desired confidence for a specific judgment task." (p.200) Therefore, if this discrepancy is large, systematic processing is more likely to be employed, whereas if it is small, heuristic processing is most likely to be preferred.

This may explain the findings of Pennycook, Epstein, et al. (2021), who observed that participants were less likely to believe misinformation when prompted to consider accuracy. Per the sufficiency principle, this prompt likely increased their desired confidence for the task at hand, which encouraged them to engage in systematic processing. Another study likewise found that when given more time to deliberate, participants were less likely to believe false social media headlines compared to when they were rushed and less attentive (Bago et al., 2020). Participants also corrected intuitive mistakes when allowed to deliberate.

The second assumption of the HSM concerns the three types of motivation that guide information processing. The first type is *accuracy motivation*, when individuals want to make well-informed judgements on the information they are processing. This motivation usually leads to systematic processing, though people may still rely on heuristics if they feel sufficiently confident in their judgment. This further explains the findings of Pennycook et al. (2021), in which the prompt participants were exposed to seems to have motivated them to seek accuracy.

The second type is *defense motivation*, in which individuals seek to protect identity-related beliefs. Although they may engage in systematic processing, it is often biased towards reinforcing their initial worldview or discrediting opposing information. Defense motivation aligns with the construct of *motivated reasoning,* commonly defined as "[t]he goal of protecting one's identity or standing in an affinity group that shares fundamental values" (Kahan, 2013, p.408). Kahan (2013) further demonstrated that heuristic processing is not the only processing mode responsible for defense motivation: participants with the highest Cognitive Reflection Test (CRT) scores were, in fact, the most prone to ideologically motivated reasoning.

However, these findings have been contested, with some studies suggesting that individuals who engage in more analytical reasoning are less susceptible to motivated reasoning (Pennycook,

Bear, et al., 2020; Pennycook & Rand, 2019). As discussed in section 2.1.1, Pennycook & Rand (2019) found that individuals who scored higher on the CRT were less likely to judge fake news as accurate and more likely to correctly identify real headlines. This was the case regardless of political ideology or headline partisanship. The authors concluded that motivated reasoning is not the primary factor in discerning fake news from real news, rather, the propensity and ability to engage in analytical thinking are more important.

The third type of motivation is *impression motivation*, where individuals aim to present themselves in a favorable light by evaluating the social acceptability of the information they are processing. This motivation can lead individuals to engage in either systematic or heuristic processing, depending on which one better serves the goal of leaving a good impression. For example, social media users may share stories that align with their peer values, not because they believe the information is accurate, but because it signals membership to the group.

These three types of motivation, together with the sufficiency principle, unconsciously influence which processing mode is engaged. However, the HSM emphasizes that heuristic and systematic processing are not mutually exclusive. Rather, they can occur simultaneously. This interaction is explained through three hypotheses: bias, additivity and attenuation (Todorov et al., 2002).

The *bias hypothesis* states that when both heuristic and systematic processing happen together, heuristic cues can skew or bias the outcome of systematic analysis. The *additivity hypothesis* posits that if both modes are engaged and point towards the same conclusion, the persuasiveness of the message will be even stronger. Finally, the *attenuation hypothesis* suggests that when the conclusions drawn from heuristic and systematic processing are in conflict, systematic processing can weaken the impact of heuristic processing.

Together, these hypotheses highlight the complex nature of information processing. Alongside motivation type, they also confirm that isolating systematic processing is not necessarily required, nor possible, to prevent disinformation from being credible and may even work against

that goal. Instead, interventions should aim to promote accuracy motivation and raise people's desired confidence levels, thus justifying using more cognitive resources.

2.3 The Dissemination of Disinformation

Multiple mechanisms enable the fast and effective spread of misinformation and disinformation, such as algorithmic amplification, the use of bots, and promotion by influential figures (Di Domenico et al., 2021). This thesis focuses on social media engagement as a key instrument in the dissemination and legitimization of disinformation.

In marketing research, the construct of social media engagement is closely related to customer engagement. It is generally conceptualized as a positive construct, representing interactions that customers have with a brand, its community, or its network (Trunfio & Rossi, 2021). Beyond its behavioral dimension, engagement can also include cognitive and affective elements, reflecting a consumer's mental and emotional connection to the object of engagement (Syrdal & Briggs, 2018). In this thesis, engagement is defined as the behavioral acts of interacting with content after it has been consumed, specifically through the actions of sharing, reacting, commenting, and saving. The construct measured is willingness to engage, which refers to the self-reported likelihood of performing these behaviors.

2.3.1 Social Media Engagement

Social media has significantly lowered the barriers to producing and spreading content. With users now central to the flow of information, they are not only passive recipients but active participants in spreading disinformation, intentionally or not. For example, two-thirds of the participants in Chadwick et al. (2018) admitted having shared problematic news content in the past month, some even knowing that the news was made up or exaggerated at the moment of sharing.

Further highlighting the diminished role of accuracy behind sharing, Pennycook, Epstein, et al. (2021) found that political alignment played a stronger role than accuracy in predicting sharing

intentions. Indeed, participants were 19.3% more likely to share politically concordant headlines rather than discordant ones, compared to only a 5.9% increase for accurate over inaccurate headlines. A notable proportion of participants indicated a willingness to share politically aligned headlines even though they did not perceive them as accurate, highlighting a disconnect between accuracy and sharing.

Among engagement behaviors, sharing appears to be the most studied (e.g., Chadwick et al., 2018), likely because it is a high-effort and public action (Molina et al., 2023). Sundar et al. (2025) describe sharing as "a way of externalizing information to others, signaling that the information provider is confident and welcoming of future discussion with others in the shared news environment. (p.161)" As such, sharing functions as the primary mechanism for organic dissemination, enabling users to amplify content from their own network or from public sources. Depending on the platform, users can add their own commentary to the shared post, further engaging with it. Sharing can also occur privately, through direct messages.

Other forms of engagement include *reacting* (i.e. "liking" or leaving emojis), *commenting*, and *saving* the content for later, all of which vary in visibility and effort. For instance, commenting, while more cognitively demanding, can allow for more nuanced interaction: one can disagree with the content, seek to provide additional information, or show support for it. Alternatively, saving may signal a will to process it more systematically later. Interestingly, Molina et al. (2023) found that false articles were more likely to receive comments, while real articles received more likes, suggesting different engagement patterns based on content accuracy.

Since user attention is limited, content is often designed to be visually and emotionally compelling, with the goal of maximizing reach, commonly known as *going viral*. Research suggests that, in addition to the network homogeneity, emotions and arousal, rather than information quality, are important predictors of virality (Lewandowsky et al., 2017). In fact, a quantitative analysis of Facebook data by Del Vicario et al. (2015) shows that most virality is driven primarily by the *echo chamber effect* and typically occurs within the first two hours after content is published, leaving little time for users to engage in a systematic analysis of the

information. Similarly, Chadwick et al. (2018) observed that the more politically homogeneous one's network is, the less likely people get challenged when they share fake news, reinforcing the behavior over time.

Additionally, users often engage with content without fully consuming it. Pennycook, Binnendyk, et al. (2021) found that individuals on social media tend to focus on headlines rather than read full articles. Supporting this, Sundar et al. (2025), in an analysis of 35 million public Facebook posts, found that approximately 75% of shared links were not clicked on before being reshared. This behavior was even more pronounced for politically aligned content, where conservatives were more likely to share false news (fact-checked and tagged as such) without clicking, while liberals were more likely to share politically aligned fact-checked real news without clicking. The authors note that this discrepancy might also be due to the flagged false content coming disproportionately from conservative sources.

Engagement also becomes a way to legitimize disinformation, regardless of user intention. Di Domenico et al. (2022) define legitimization as the multidimensional process of harnessing and gathering socio-cultural support for the ideas behind misleading information, which in turn confers credibility. One form of this is algorithmic legitimacy, where platform algorithms, driven by engagement metrics, amplify the visibility of disinformation. These metrics act as credibility heuristic cues and shape users' perception of how widespread or accepted certain ideas are (Pang et al., 2016). This process can be further manipulated through the use of bots and astroturfing, artificially inflating engagement and creating a false sense of consensus (Molina et al., 2023).

In summary, while individuals may engage with false information for various reasons, such actions often contribute to spreading it further and legitimizing it. This is amplified by algorithms that prioritize content with a high engagement and repeatedly promote topics users have previously interacted with, reinforcing informational echo chambers. Moreover, the prevalence of superficial engagement, such as sharing without reading, prevents people from processing information systematically and encourages heuristic processing driven by emotions

and biases. This highlights the need for multidimensional interventions that disrupt engagement at both the algorithmic and individual levels.

2.4 Warning Interventions and Disinformation

Building on the previous sections, one promising strategy to reduce the credibility and spread of disinformation is preemptive intervention. According to *inoculation theory*, exposing individuals to a weakened form of persuasive misinformation can build resistance to it, much like a vaccine against viruses (Lewandowsky & van der Linden, 2021). This process, known as *prebunking,* typically includes two elements: a general warning about the potential presence of disinformation, and an explanation of the deceptive techniques or logical fallacies being used.

Warnings have been used in both academic research and practical applications to reduce the spread and belief of false information. In this research, warnings are defined as a form of information disclosure presented as a visual indicator and/or message, designed to draw attention to a specific aspect of the content that warrants attention (Mende et al., 2024). However, their effectiveness has been debated. One concern is the familiarity backfire effect, where repeating false information in the process of correcting it may paradoxically increase belief in the misinformation (Ecker et al., 2020). For instance, Nyhan & Reifler (2010) found evidence of this effect in politically charged contexts. When participants were shown corrections related to controversial topics, such as the presence of weapons of mass destruction in Iraq or the true costs of tax cuts, those whose views were discordant not only resisted the correction, but, in some cases, strengthened their original beliefs. The authors attribute this response to motivated reasoning.

Yet, other studies challenge the existence of the backfire effect. Pennycook et al. (2020) found that warnings actually reduced belief in false headlines, even when the content was politically discordant. In fact, warnings were more effective for politically discordant headlines than for concordant ones. Similarly, Ecker et al. (2020) found mixed evidence of the backfire effect: while one experiment suggested that exposure to fact-checked information alone could increase

belief in the false information, follow-up experiments did not replicate this effect. Overall, the authors concluded that while familiarity may interfere with correction, it is unlikely to produce a significant backfire effect.

Despite these concerns, a growing body of research supports using warnings as a solution to decrease disinformation credibility (e.g., Hameleers et al., 2020; Lee & Shin, 2022; Pennycook et al., 2020). For example, Hameleers et al. (2020) found that fact-checks, whether visual or textual, effectively reduced credibility ratings for both text-only and text-with-image disinformation, regardless of the format used for correction.

Importantly, warnings may also influence engagement with disinformation content. For example, Lee & Shin (2022) showed that false tags attached to fake news posts significantly reduced intent to engage, especially when the content involved highly vivid sources such as deepfake videos. The false-tags likely acted as nudges, triggering accuracy-motivated information processing. Similarly, Pennycook et al. (2020) observed that false headlines labeled with a warning were less likely to be considered for sharing. While other research suggests that sharing behavior is not always tied to accuracy judgments (e.g., Pennycook et al., 2021), prompting individuals to consider accuracy before sharing has been shown to improve the overall quality of shared content.

From the perspective of the Heuristic-Systematic Model (HSM), warnings may work by increasing the discrepancy between individuals' actual and desired confidence (via the sufficiency principle). This discrepancy can prompt individuals to be more vigilant about the content they encounter, encouraging a shift to more systematic processing and activating accuracy motivation.

Although the HSM does not guarantee that systematic processing will lead to the correct identification of disinformation, it is likely to increase scrutiny, which can reduce both the willingness to engage with and the credibility of disinformation.

Based on this logic, the following hypothesis is proposed:

- **H1:** Exposure to a warning will decrease participants' willingness to engage with disinformation content.

2.4.1 Types of Warnings

Warnings used to counter disinformation vary widely in form, presentation, and content. This variability has prompted research into what elements make warnings more effective. For instance, studies suggest that warnings are more impactful when they target a specific content piece rather than deliver a general alert, and when they are highly visible to users (Martel & Rand, 2023). Their effectiveness further increases when they disrupt the interaction flow, for example, by requiring users to close a modal window before accessing the content (Mende et al., 2024).

The specificity of warning content is also important. Ideally, warnings should provide relevant corrective information to maximize their effectiveness (Lewandowsky et al., 2017). However, this approach is often impractical at scale, as the spread of disinformation outpaces the ability of fact-checkers to provide accurate corrections. This raises concerns about the *implied truth effect*, in which inconsistently flagged content leads users to perceive unflagged information as more credible (Pennycook, Bear, et al., 2020).

Finally, the perceived credibility of the warning source can influence its effectiveness. As Booth et al. (2024) point out, even accurate warnings may be ineffective or even counterproductive if the source is viewed as untrustworthy. For example, conspiracy-minded individuals may dismiss fact checks issued by governmental institutions or mainstream media outlets, regardless of the accuracy, because they do not trust the source.

In summary, effective warning labels tend to be visible, placed adjacent to the content, interruptive, and specific. They should also come from a source that is seen as neutral and credible by the target audience to be effectively inoculating. However, given the challenges of

providing specific and timely corrections, this study employs a generalized warning that is brief, visible, and positioned just before the exposure to misleading content. While such warnings may be less powerful than highly specific ones to inoculate against disinformation, prior work suggests they can still reduce disinformation credibility and willingness to engage.

2.5 Engagement and Credibility

Credibility has often been studied through the lens of source credibility, where the perceived trustworthiness and expertise of the source influence the evaluation of information credibility (Metzger & Flanagin, 2013). In information science, however, the focus has mostly been on the credibility of the message itself, which is particularly relevant in the context of online media where the source is often ambiguous or unknown. According to Appelman and Sundar (2016), message credibility refers to "an individual's judgment of the veracity of the content of communication" (p.63). This study adopts this definition, operationalizing credibility as perceived credibility, as it is self-reported by participants.

Metzger & Flanagin (2013) argue that although credibility evaluations should be done through systematic processing, users typically rely on heuristics rather than deep evaluation when assessing online content. This reliance on heuristics is central to Sundar's MAIN model, which explains how media affordances (e.g., likes, shares, modality) provide cues that trigger credibility-related heuristics (Sundar, 2008). According to the model, every digital platform is embedded with affordances that guide how users interact with the platform and shape the platform itself. More precisely, these affordances act as a library of cues that inform how information is heuristically processed and help determine the credibility of the content. For example, engagement metrics, like share counts or comments, activate the *bandwagon heuristic*, leading users to assess content credibility or quality based on its popularity. For a heuristic to be used, it must be readily accessible and relevant to the task at hand.

Empirical evidence supports the role of these cues. For instance, Metzger et al. (2010) found that people often rely on engagement signals, such as endorsement and user ratings, to inform their

credibility judgments. Colliander (2019) showed that negative user comments can decrease favorable attitudes and reduced sharing intentions, suggesting that the outcomes of the bandwagon effect can be positive or negative.

While prior studies suggest that people may sometimes share content they know or suspect to be inaccurate (e.g., Chadwick et al., 2018; Molina et al., 2023; Pennycook et al., 2020), credibility still appears to influence engagement, even if indirectly. For example, Sundar et al. (2021) found that videos were perceived as both more credible and more likely to be shared than other modalities, implying a positive relationship between credibility and engagement (although this relationship was not tested directly). Similarly, Luo et al. (2022) found that a high number of Facebook likes increased message credibility while simultaneously reducing participants' ability to accurately detect fake news, demonstrating the persuasiveness of the bandwagon heuristic.

The type of engagement also appears to matter when it comes to credibility. Molina et al. (2023) observed that users had higher intentions to comment on false posts because the content made them feel uneasy. Yet, the authors observed that these differences were not present when users read the full article rather than only the social media post. Likewise, Metzger et al. (2021) reported that users may share questionable content to crowdsource its credibility assessment to their network or to educate their connections through sarcasm or mockery.

Overall, while the influence of credibility seems to vary depending on the engagement behaviors, evidence suggests that more credible content generally elicits higher engagement. Considering that, as noted earlier, warnings tend to lower credibility evaluations, their presence may indirectly reduce the intention to engage by first undermining perceived credibility and then decreasing willingness to interact with the content.

This leads us to the following mediation hypothesis:

- **H2:** Perceived credibility will mediate the effect of warnings on engagement, such that exposure to a warning reduces credibility, which in turn reduces willingness to engage.

2.6 Disinformation and Multimodality

Disinformation has been flourishing on social media, where platform affordances enable widespread and rapid dissemination. One of those key affordances is the ability to engage across multiple content modalities. A modality refers to a method of information presentation and processing associated with one of the five senses (Fisher et al., 2019), such as audio for auditory processing. Multimodality describes the presence of more than one mode of communication within a message or environment, for example, combining text, images, audio, and video (Sundar, 2008). In this sense, social media is multimodal as it exposes users to a variety of content formats.

This section explores how content modality may influence the credibility of disinformation.

2.6.1 The Moderating Role of Multimodality

The rise of multimodal content has introduced an additional layer of complexity to how people engage with information online. Compared to early text-based internet content, the ability to create, share, and visualize information in multiple formats has opened up new avenues for creators of disinformation.

According to the Heuristic-Systematic Model (HSM), individuals tend to rely on heuristic processing when motivation or the ability to process information systematically are low. In line with this, Lang (2006) argues that real-time formats, such as audiovisual content on television or audio-only content on radio, require more cognitive resources to encode than text because they contain peripheral cues that need to be processed in addition to the message itself. This additional cognitive load can lower the ability to process the message systematically, increasing reliance on heuristics cues, which, in turn, can increase perceived credibility.

This idea is expanded by Sundar's MAIN model (Sundar, 2008), which confirms that modality is a digital affordance that triggers various credibility-related heuristic cues. Sundar explains that "There are three possible origins of cognitive heuristics within this affordance: (1) each individual modality (e.g., text, aural, audiovisual) may, by its sheer presence, cue a particular heuristic; (2) new modalities unique to digital media could also cue their own heuristics; and (3) combinations of modalities may cue heuristics as well." (p.80). Consistent with Lang's (2006) argument, the primary heuristic associated with modality is the *realism heuristic*, which suggests that the more realistic and sensory-rich the medium, the more credible it appears.

The Media Richness Theory (Daft & Lengel, 1986) offers an additional framework for interpreting modality effects. Originally developed to guide communication strategies in organizations, the MRT defines *richness* as a medium's ability to convey multiple cues and support rapid feedback, thus reducing ambiguity and uncertainty. Richer media (e.g., video) allow for greater range of verbal and nonverbal information, including tone, facial expressions, and context cues, while leaner media (e.g., text) are more suitable for straightforward, unambiguous content. In the context of disinformation, rich media may appear more trustworthy because they convey more contextual cues that reduce uncertainty in the interpretation of the message. This interpretation aligns closely with the MAIN model's realism heuristic, offering further support for the notion that modality cues may shape credibility assessments.

Recent studies have provided evidence in support of this idea. For instance, Sundar et al. (2021) found that audiovisual content was perceived as more credible than text-only and audio-only content, largely due to the perceived realism of videos. This increase also led to a higher intent to share content. Interestingly, audio-only content was rated as more credible than text, showing a spectrum of credibility based on modality. In line with the HSM's additivity hypothesis, where systematic and heuristic processing work together to strengthen the persuasiveness of the information, the authors argue that the realism heuristic can engage both processes simultaneously.

Further supporting this, Yadav et al. (2011), found that videos were perceived as more engaging than text for the same content, likely due to their vivid realism, although participants retained information as effectively for both modalities. Likewise, Smelter & Calvillo (2020) demonstrated that semantically appropriate images increased the perceived truthfulness of misinformation compared to text-only versions. This was corroborated by Lee & Shin (2022) who found that deepfakes were perceived as the most salient and credible, followed by text with images, both of which lead to higher credibility levels than text alone. Conversely, Hameleers et al. (2020) found only partial evidence for a modality effect on credibility: while text with images was seen as more credible than text-only for disinformation about refugees, this pattern did not hold for content about school shootings. Despite some topic-specific variations, visuals overall appear to help with memory retrieval and are generally perceived as more credible due to their realism or vividness (Vaccari & Chadwick, 2020).

Taken together, these findings suggest that content modality shapes how disinformation is perceived, both by triggering specific heuristics and by reducing the cognitive resources available for systematic processing. Since different modalities seem to elicit varying levels of perceived credibility, it is likely that the effectiveness of warnings will also vary depending on content format, with richer media like video being seen as more credible and therefore less affected by warnings. This leads us to the following hypothesis:

- **H3:** Content modality will moderate the effect of warning exposure on perceived credibility, such that warnings will reduce credibility less in richer modalities (e.g., video).

2.7 Conceptual Model

Figure 1 summarizes the conceptual model developed from the literature review. The hypotheses outlined throughout this section are visually integrated here to clarify the suggested relationships

between warnings, modality, credibility, and engagement. This model serves as the foundation for the experimental design presented in the next chapter.

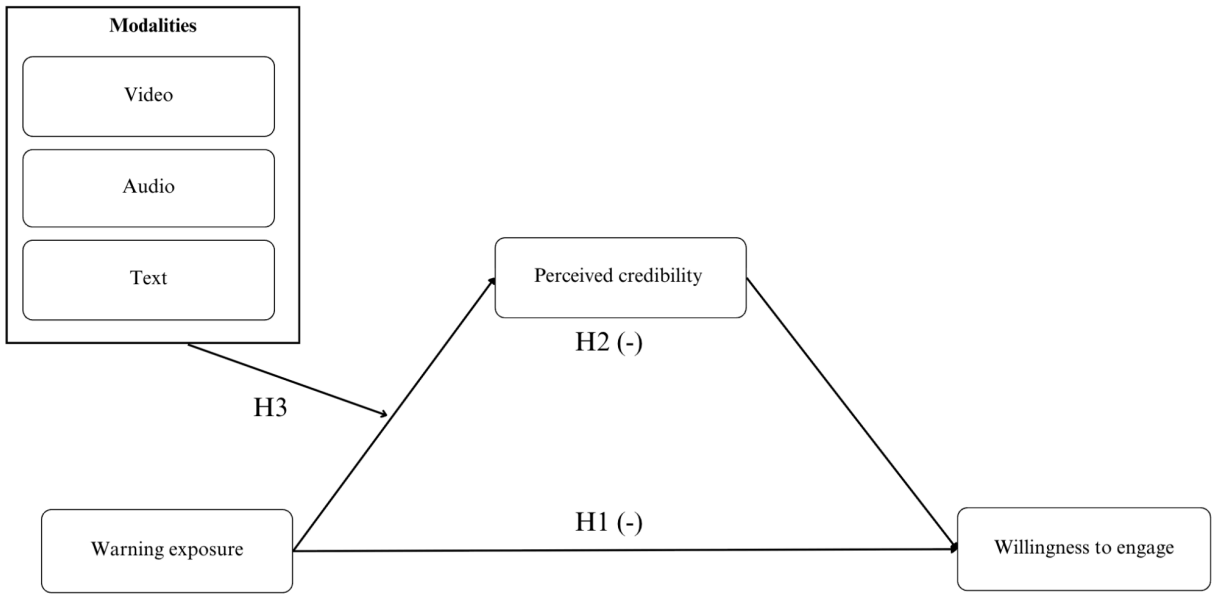


Figure 1. *Conceptual model*

Chapter 3
Methodology

This chapter describes the methodology of this research in detail. It outlines the process for both the pretest and the main experiment. It also explains the selection of variables, specifically the measures chosen for the dependent variables, as well as the different conditions within the independent variables.

3.1 Pretest

A pretest was conducted to select the specific piece of content that would be adapted for use across all modalities and to verify that the warning is sufficiently noticeable.

3.1.1 Experimental Design

A one-factor between-subjects experimental design was employed to determine which disinformation video would be used in the main experiment. The independent variable was video content, with participants randomly assigned to watch one of four disinformation videos on climate change.

The videos were sourced from YouTube and shortened to approximately two minutes to balance participant attention with sufficient content for evaluation. YouTube channels were selected using resources like DeSmog (DeSmog Climate Disinformation Database, n.d.-b), which identifies organizations and individuals known to promote climate change denial or skepticism. Notable sources included the Heartland institute, the CO_2 Coalition, and the Hoover institution, each of which maintains an active online presence. From these channels, four videos were selected based on similar production quality, and their promotion of scientifically debunked information.

The four videos are labeled as follow:

- Video 1: Glacier Retreat. Argues that climate change has no effect on glaciers and rising sea waters.
- Video 2: CO_2 and Food Abundance. Present CO_2 as a net positive for the planet, especially in large quantities.
- Video 3: Debunking Hurricane Myths. Denies that climate change increases the strength and frequencies of hurricanes.
- Video 4: The Challenges and Realities of Climate Modelling with Steven Koonin. Claims that climate models are inaccurate.

The experiment was conducted through an online survey built on Qualtrics, and distributed through Prolific, allowing participants to complete the study remotely at their own pace.

3.1.2 Sample

The initial sample included 200 participants, who were divided into four groups of 50. Participants were recruited anonymously through a convenience sample on Prolific and were paid approximately 2$ for their participation. They were screened based on whether they live in North America, their ability to fully engage with the video content, both physically and technologically, and their ability to speak English. A total of seven participants' data were excluded, which led to a final sample size of 193: four for failing attention checks, and three for failing comprehension checks.

The most common age group was 25-34 years old (34%), followed by 35-44 years old (22%), and 18-24 years old (19%). Twenty-three percent of participants were over 45 years old, among whom only five identified as older than 64. Fifty-eight percent of participants identified as female, 38% as male, and the remainder as non-binary.

3.1.3 Procedure

First, a warning was presented to all participants below the instructions for the study. Participants were then randomly assigned to one of the four video conditions and, after viewing the video,

answered a comprehension question to ensure they had watched enough of the content. They then rated the credibility of the video, followed by their willingness to engage with it. Next, a manipulation check for the warning condition was conducted by asking participants whether they had noticed the warning message. To control for potential confounding effects related to video quality, participants also rated the video on interest, video production, audio production, and content organization. They were also asked if they had previously encountered the information presented in the video.

Then, participants completed a brief demographic questionnaire (age range, gender, political leaning, and education), followed by questions about their environmental attitude. The full questionnaire is available in **Appendix A**.

Finally, a thorough debrief (see **Appendix C**) was presented at the end to explain the study's complete purpose and provide resources to help counter the disinformation they had been exposed to.

3.1.4 Measures

Participants were asked to evaluate one of four videos using both a subset of the measures later used in the main study, as well as additional measures designed to support the selection of the final stimulus.

Perceived credibility was measured using a seven-point Likert scale from "Very Poorly" to "Very well") based on a scale developed by Appelman & Sundar (2016). Participants rated how well the following adjectives described the content: accuracy, authenticity, and believability (three items, $M = 4.61$, $SD = 1.77$, $\alpha = .92$).

Additionally, we measured *willingness to engage* with the content as an attitudinal indicator. This measure is essential because the more frequently content is engaged with, the more disinformation spreads and persists (Molina et al., 2023). Following a similar structure to Pang et al. (2016), engagement was assessed by asking participants to rate the likelihood they were to

share, comment on, or react to the content they were exposed to, using a five-point Likert scale (from “Extremely unlikely” to “Extremely likely”; three items; $M = 2.30$, $SD = 1.34$, $\alpha = .85$).

Video quality perception was evaluated using three five-point Likert scales items (from “Terrible” to “Excellent”) assessing video production quality, audio production quality, and content organization ($M = 3.92$, $SD = 0.92$, $\alpha = .77$). Interest was also controlled for using a five-point Likert scale ranging from “Not interesting at all” to “Extremely interesting”. Prior exposure to the content was measured with a yes/no/maybe item asking whether participants had previously encountered the information presented in the video.

Finally, demographic information was collected, including age range, gender, political leaning, and educational level. Environmental attitude was measured using six items of a seven-point Likert scale developed by Milfont & Duckitt (2010), as all four videos focused on climate change related topics ($M = 4.85$, $SD = 1.50$, $\alpha = .74$). This measure was included to help identify any underlying differences across groups that might influence how participants evaluate the disinformation content due to prior attitudes toward the environment.

3.1.5 Results

The pretest results showed that most videos were rated similarly in terms of production quality (see **Table 1**), except for Video 4 ($M = 3.50$, $SD = 0.88$), which was rated significantly lower than the others. A one-way ANOVA followed by Games-Howell post hoc tests (used due to unequal variances and sample sizes) confirmed that Video 4 was rated significantly lower in quality than Video 1, 2, and 3 (all p's $< .01$), while no significant differences were found among the other three.

In terms of interest, although mean scores varied slightly (see **Table 1**), a Welch ANOVA indicated that these differences were not statistically significant, $F(3, 104.34) = 1.11$, $p = .348$. This suggests that none of the videos were consistently more or less interesting across participants.

Interestingly, Video 2 demonstrated the highest variability in participant response across interest (M = 2.96, SD = 1.34), credibility (M = 4.16, SD = 2.01), and engagement (M = 2.29, SD = 1.21), suggesting it may have been the most polarizing or controversial of all four stimuli. While not the most credible or engaging on average, Video 2's greater variance, combined with its comparable quality and interest level, made it a strong candidate for the main experiment, where detecting differences in credibility and engagement is critical.

The manipulation checks for the warning condition confirmed that most participants noticed it (M = 77.73%).

Table 1

Descriptive Statistics of Pretest Results

Stimuli		Interest		Quality		Credibility		Engagement	
	n	M	SD	M	SD	M	SD	M	SD
Video 1: Glaciers	49	3.04	0.93	4.05	0.62	5.29	1.33	2.54	1.18
Video 2: CO_2	49	2.96	1.34	4.05	0.73	4.16	2.01	2.29	1.21
Video 3: Hurricanes	48	2.92	1.07	4.08	0.65	4.39	1.70	2.22	1.18
Video 4: Models	47	2.68	1.02	3.50	0.88	4.62	1.19	2.12	1.14

Note. n = number of participants per condition; M = mean; SD = standard deviation

Video 1: *Glaciers Retreat* by the CO_2 Coalition

Video 2: *CO_2 and Food Abundance* by the CO_2 Coalition

Video 3: *Debunking Hurricane Myths* by the Heartland Institute

Video 4: *The Challenges and Realities of Climate Modeling* by Hoover Institution

3.2 Main Study

3.2.1 Experimental Design

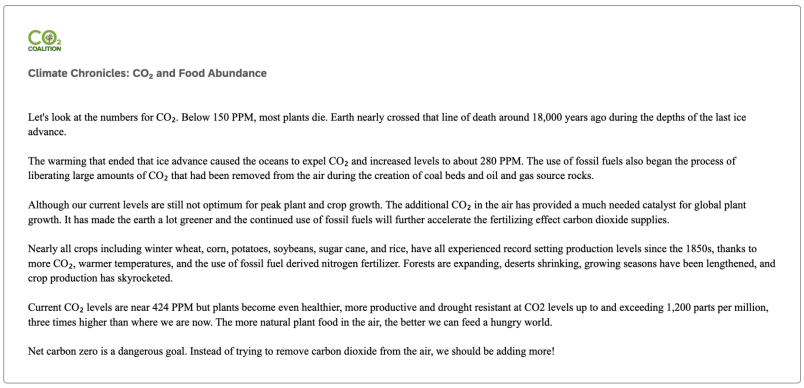
To test the proposed hypotheses, a three (modality: video, text, audio) by two (warning: presence, absence) between-subjects experiment was conducted through an online survey. Conducting the experiment online offered two advantages: it allowed the recruitment of a larger participant pool and created an experience closer to how users naturally encounter disinformation. Each participant was randomly assigned to one of the six conditions and only exposed to a single content format.

The video selected for the main study focused on the topic of CO_2 production, arguing, through dubious evidence, that CO_2 is beneficial to the environment and that the planet needs more of it in its atmosphere. The video was created by the CO_2 Coalition, a nonprofit organization known for disputing the scientific consensus on global warming, particularly the role of fossil fuel and CO_2 (DeSmog Climate Disinformation Database, n.d.-a). The original four-minute video was trimmed to two minutes to maintain participant attention. It featured the voice of a British female narrator, stock videos, and a few charts. While people appear in the stock footage, they are not the focus of the video.

The conditions for the modality factor were chosen based on the three content formats most commonly encountered on social media: video, audio, and text. To maintain consistency across conditions, all versions included the original video's source (which could not be removed), since prior research shows that source information can significantly influence credibility perceptions (Ali et al., 2021; Metzger & Flanagin, 2013).

To match the professional-looking format of the video condition (see **Figure 2**), which appeared in a dedicated video player on Qualtrics, the text condition was designed to resemble an article (see **Figure 3**), and the audio condition was presented in a dedicated audio player (see **Figure 4**).

Figure 2. *Video modality*



Climate Chronicles: CO_2 and Food Abundance

Let's look at the numbers for CO_2. Below 150 PPM, most plants die. Earth nearly crossed that line of death around 18,000 years ago during the depths of the last ice advance.

The warming that ended that ice advance caused the oceans to expel CO_2 and increased levels to about 280 PPM. The use of fossil fuels also began the process of liberating large amounts of CO_2 that had been removed from the air during the creation of coal beds and oil and gas source rocks.

Although our current levels are still not optimum for peak plant and crop growth. The additional CO_2 in the air has provided a much needed catalyst for global plant growth. It has made the earth a lot greener and the continued use of fossil fuels will further accelerate the fertilizing effect carbon dioxide supplies.

Nearly all crops including winter wheat, corn, potatoes, soybeans, sugar cane, and rice, have all experienced record setting production levels since the 1850s, thanks to more CO_2, warmer temperatures, and the use of fossil fuel derived nitrogen fertilizer. Forests are expanding, deserts shrinking, growing seasons have been lengthened, and crop production has skyrocketed.

Current CO_2 levels are near 424 PPM but plants become even healthier, more productive and drought resistant at CO2 levels up to and exceeding 1,200 parts per million, three times higher than where we are now. The more natural plant food in the air, the better we can feed a hungry world.

Net carbon zero is a dangerous goal. Instead of trying to remove carbon dioxide from the air, we should be adding more!

Figure 3. *Text modality*

Figure 4. *Audio modality*

Participants in the warning condition saw a generic warning placed directly under the instructions, on a separate page preceding the stimulus (see **Figure 5**). The warning was designed to be minimally intrusive yet noticeable. To introduce friction, participants were required to actively confirm they had read the instructions before clicking to proceed.

This approach aligns with prior research that suggests that warnings are more effective when they interrupt the interaction with the content, for instance, by forcing users to click a button to close a modal window (Mende et al., 2024). While our implementation was less intrusive, requiring interaction before proceeding added a moderate level of disruption.

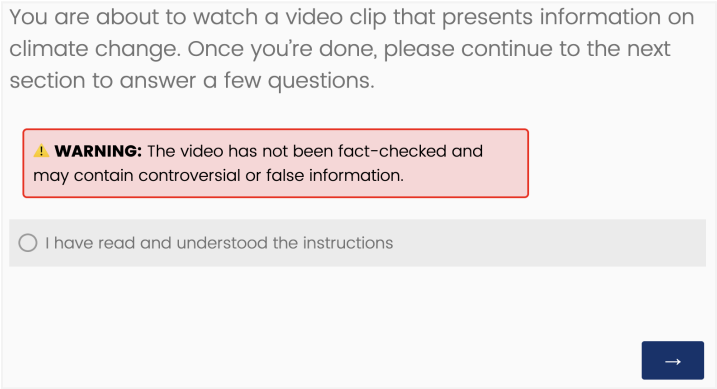


Figure 5. *Warning and instructions (video version)*

The warning itself was intentionally vague. Rather than correcting facts, or explaining disinformation tactics, it simply alerted participants that the upcoming content could contain false or misleading information. This design choice enabled us to test whether a low-effort warning could successfully be implemented at scale without requiring too many additional fact-checking resources. It also aimed to preserve participants' agency, allowing them to determine for themselves the credibility of the content.

3.2.2 Context: Climate-Change Disinformation

Climate change was selected as a topic of disinformation in this study due to its strong presence in disinformation campaigns, high public relevance, and well-documented complexity. Over the

past decades, climate change has become one of the most politicized scientific issues, making it particularly susceptible to disinformation campaigns (Tam & Chan, 2023). It is part of a broader ideological current that seeks to delegitimize science, often framing scientific research as the product of corrupt elites motivated by self-interest (Hameleers & Meer, 2021).

Climate skepticism and denial have been increasingly incorporated into the political rhetoric of populist leaders, further reinforcing public distrust in scientific institutions. For example, in the United States, the issue is deeply polarizing: Democrats are more likely to acknowledge that climate change is anthropogenic, while Republicans are more likely to deny it, a position which is reinforced by party leaders (Kahan, 2013; Lewandowsky, 2021). It is important to note that, even among climate change skeptics, the majority do not believe that climate change as a whole is a conspiracy (Tam & Chan, 2023).

In addition, multiple bad actors have actively produced and disseminated climate-change disinformation, often for economic or political reasons. Perhaps the most well-known case is the oil industry having invested for decades in messages and studies designed to challenge climate science in order to continue the exploitation of fossil fuels (Pierre & Neuman, 2021). An example of this disinformation is the creation of the concept of *carbon footprint,* developed by British Petroleum (BP) in association with a marketing agency to shift responsibility for climate change from corporations to individuals (Kaufman, 2020). Similarly, conservative think tanks have played a major role in producing and promoting climate change skepticism. Lewandowsky (2021) notes, for instance, that more than 90% of books promoting climate skepticism and denial are sponsored by such organizations.

Common rhetorical strategies used in climate change disinformation include the following (Diaz Ruiz & Nilsson, 2023; Lewandowsky, 2021; Meta, 2022):

- **Impossible expectations**: Demanding unreasonable levels of certainty from climate scientists, and in doing so, casting systematic doubt in science, and highlighting uncertainties as proof that the consensus is wrong.

- **Cherry picking**: Highlighting carefully selected data out of context or isolated anecdotes (e.g., a particularly cold winter) to contradict broader climate trends.
- **Single-cause fallacy**: Attributing complex phenomena to a single cause. For example, blaming solar activity for rising temperatures.
- **Fake expert and pseudoscience**: Painting experts who agree with the scientific consensus on climate change as biased or corrupt while elevating fake experts as credible figures fighting to expose the truth.
- **False equivalence**: Presenting two different situations as equivalent, such as suggesting past weather anomalies are evidence that current climate disasters are expected.
- **Undermining institutions and scientists**: Using ad hominem attacks on scientists, questioning motivations, or discrediting scientific methodology (e.g., peer reviews).

Beyond its political relevance and long history of disinformation, climate change was selected for the present experiment for the following reasons. First, it is a widely known issue, ensuring baseline familiarity across participants. Second, its scientific complexity makes it difficult for laypeople to independently verify information, which may impact their ability to engage in systematic processing. Third, climate change has been the target of extensive and well-documented disinformation campaigns, making it a strong candidate for testing the efficacy of warnings. Finally, climate change related content can easily be found in multiple formats on social media, justifying and facilitating the use of multimodal content. In summary, climate change disinformation provides an ecologically valid context for testing how content modality and warning interventions influence credibility and engagement.

3.2.3 Sample

The total sample size was 570, with approximately 95 participants per condition. After exclusions, 553 participants remained, resulting in slightly uneven group sizes (see **Table 2**).

Participants' data were excluded if they failed the attention or comprehension check, spent less than 60 seconds on the video or audio page, or less than 25 seconds on the text page. Page viewing times were used as a proxy for whether participants had sufficiently engaged with the content. The thresholds for time were set based on content length and expected minimum engagement. Text was treated with more flexibility due to variance in reading speed.

Table 2

Sample Size by Experimental Condition

Modality	Warning	n
Video	Present	93
	Absent	95
Text	Present	95
	Absent	91
Audio	Present	89
	Absent	90

A convenience sampling method was used. Participants were recruited anonymously through Prolific and were paid approximately 2$ for their participation. The inclusion criteria were the following: participants must be over 18 years old, reside in Canada or the United States, be able to engage with the content fully (i.e., without a sensory disability preventing it), understand English, and have access to a suitable electronic device.

Most participants identified as female (57.7%), followed by male (41.7%), and non-binary or other (0.6%). Five participants declined to share their gender. Participants aged 18-34 made up 41.5% of the sample, those aged 35-54 accounted for 44.6%, and 13.9% were aged 55 or older. Two participants did not specify their age range. In terms of education, most participants had an

undergraduate or professional degree (44.2%), followed by some college education or an associate degree (26.3%), a graduate degree (16.7%), a high school diploma (12%), or less than a high school degree (1%). Two participants did not answer the education question. Politically, 41% identified as left-leaning to some degree, 29.6% as centrist, and 29.4% as right-leaning to some degree. Thirteen participants did not specify a political affiliation.

3.2.4 Procedure

To start, participants were randomly assigned to one of three modality groups (text, audio, or video), and each group was further divided based on whether they would be exposed to a warning prior to viewing the content. The warning, when present, appeared right below the instruction text specific to the modality (see **Figure 5**), and participants were required to click a button to proceed to the stimulus page. All participants were provided with the same instructions, tailored to their assigned modality.

After viewing the content, participants answered a comprehension question to ensure they had processed the material adequately. Those who answered incorrectly were given a second opportunity to review the content and answer again. Next, participants were asked to rate the credibility of the content, followed by their willingness to engage with it. They also answered a behavioral question that verified if they wanted to receive additional information on the topic. After this, participants filled out a series of scales evaluating their environmental attitude, and science media literacy.

Finally, participants provided basic demographic information, including age range, gender, political leaning, and level of education.

After completion, participants were shown a debriefing page that explained the true purpose of the study, provided accurate information to correct the disinformation they were exposed to, and offered resources to help them identify and resist similar content in the future (see **Appendix D**). There was no time limit to complete the questionnaire, which typically took between five and ten minutes to complete.

3.2.5 Measures

The main study used the same constructs introduced in the pretest, perceived credibility and willingness to engage, to assess the impact of warnings and modality on how believable and engaging disinformation is.

Perceived credibility was measured using the same three-item scale from Appelman & Sundar (2016), where participants rated the content on accuracy, authenticity, and believability using a seven-point Likert scale ranging from "Very poorly" to "Very well" (see **Appendix E)**.

Willingness to engage was assessed through participants' likelihood to share, comment, or react to the content. These were rated on a seven-point Likert scale from "Extremely unlikely" to "Extremely likely", capturing attitudinal engagement (see **Appendix F)**.

In addition, a behavioral engagement measure asked whether participants wanted to seek more information on the topic after being exposed to the content. Although it does not capture the motivation behind the interest, whether agreement, curiosity, or skepticism, it provides an indicator of engagement that goes beyond self-reported intentions.

To account for individual differences, the study also included several control variables:

- *Demographic information* (age range, gender, and education level) was collected through multiple choice questions, with participants given the option to skip any of these three questions.
- *Political affiliation* was recorded using a seven-point Likert scale going from "Strongly left-leaning" to "Strongly right-leaning". This variable was included due to the influence of politics on disinformation processing, especially on topics like climate change (Tam & Chan, 2023). Given the sensitivity of this information, participants were allowed to leave it unanswered.
- *Environmental attitude* was assessed using six items of a seven-point Likert scale developed by Milfont & Duckitt (2010) (see **Appendix G)**. Like in the pretest, this measure was included to account for the possibility that responses to the climate change

related disinformation stimulus may be shaped by participants' existing environmental attitudes.

- *Science media literacy* was measured using four items of a six-point Likert scale developed by Austin et al. (2023), to account for participants' ability to critically assess scientific online content (see **Appendix H)**.

Finally, reliability measures for all scales used in the study are noted at the beginning of Chapter 4 and the full questionnaire, including demographic and political affiliation questions, is available in **Appendix B**.

Chapter 4
Results

This chapter presents the results of the main study. The data was analyzed using SPSS Statistics and Excel.

4.1 Measurements Reliability and Control Variables

The scales for both the key dependent variable, engagement (three items, $M = 3.03$, $SD = 1.78$, $\alpha = .865$), and the mediator, credibility (three items, $M = 3.95$, $SD = 1.83$, $\alpha = .932$, Appelman & Sundar, 2016) were found to be reliable.

Several control variables were included: environmental attitude (six items, $M = 4.78$, $SD = 1.73$, $\alpha = .743$, Milfont & Duckitt, 2010), science media literacy (four items, $M = 3.94$, $SD = 0.97$, $\alpha = .768$, Austin et al., 2023), and political affiliation (seven-point Likert scale, from "Strongly left-leaning" to "Strongly right-leaning"). Age range, education, and gender were also considered, but were excluded from the final models as they were not significant in any of the main effect tests (all $p > .10$).

4.2 Descriptive Statistics

The results highlighted certain trends for both credibility and engagement. This section presents and discusses the means of the key variables before proceeding to the hypothesis testing.

4.2.1 Credibility

The most noticeable pattern is that credibility is higher when no warning was presented to participants across modalities ($M_{warning} = 3.64$, $SD_{warning} = 1.80$; $M_{no\text{-}warning} = 4.25$, $SD_{no\text{-}warning} = 1.83$). The audio modality sees the largest increase in credibility when warning is absent (+0.70), followed by text (+0.67), and video (+0.45). The difference in scores is minor between audio and

video (M_{video} = 4.07, SD_{video} = 1.91; M_{audio} = 4.08, SD_{audio} = 1.77), and larger with text (M_{text} = 3.69, SD_{text} = 1.79). Another noteworthy pattern is that *accuracy* is the lowest item on the scale across modalities while *believability* is the highest one (see **Table 3**).

Table 3.

Descriptive Statistics for Perceived Credibility by Modality and Warning Condition

		Average per item			Total	
		Accurate	Authentic	Believable	*M*	*SD*
Video	Warning	3.60	3.83	4.08	3.84	1.86
	No warning	4.00	4.22	4.66	4.29	1.93
	Total	3.80	4.03	4.37	4.07	1.91
Text	Warning	3.28	3.38	3.43	3.36	1.72
	No warning	3.93	4.02	4.14	4.03	1.81
	Total	3.60	3.69	3.78	3.69	1.79
Audio	Warning	3.45	3.63	4.11	3.73	1.72
	No warning	4.17	4.52	4.60	4.43	1.75
	Total	3.81	4.08	4.36	4.08	1.77
Total	Warning	3.44	3.61	3.87	3.64	1.80
	No warning	4.03	4.25	4.47	4.25	1.83

4.2.2 Engagement

A similar pattern is observed with engagement, where the presence of a warning decreases engagement scores across modalities (see **Table 4**), though this effect is minimal for text ($M_{text\text{-}warning}$ = 2.85, $SD_{text\text{-}warning}$ = 1.70; $M_{text\text{-}nowarning}$ = 2.88, $SD_{text\text{-}nowarning}$ = 1.77).

Across modalities, audio records the lowest scores when a warning is present ($M_{audio\text{-}warning}$ = 2.46, $SD_{audio\text{-}warning}$ = 1.60), while text shows the lowest score when the warning is absent ($M_{text\text{-}nowarning}$ =

2.88, $SD_{\text{text-nowarning}}$ = 1.77). Despite this, the overall engagement scores for audio and text are nearly identical (M_{audio} = 2.86, SD_{audio} = 1.77; M_{text} = 2.87, SD_{text} = 1.73).

The video modality consistently receives the highest engagement scores, for both the warning and no warning conditions ($M_{\text{video-warning}}$ = 3.07, $SD_{\text{video-warning}}$ = 1.66; $M_{\text{video-nowarning}}$ = 3.64, $SD_{\text{video-nowarning}}$ = 1.91).

Table 4.

Descriptive Statistics for Willingness to Engage by Modality and Warning Condition

		Average per item			Total	
		Reaction	Share	Comment	M	SD
Video	Warning	3.40	2.60	3.22	3.07	1.66
	No warning	4.23	3.26	3.44	3.64	1.91
	Total	3.81	2.93	3.33	3.36	1.80
Text	Warning	3.17	2.44	2.94	2.85	1.70
	No warning	3.23	2.62	2.80	2.88	1.77
	Total	3.20	2.53	2.87	2.87	1.73
Audio	Warning	2.91	1.97	2.52	2.46	1.60
	No warning	3.64	2.91	3.18	3.24	1.86
	Total	3.28	2.44	2.85	2.86	1.77
Total	Warning	3.16	2.65	2.90	2.80	1.67
	No warning	3.71	2.94	3.14	3.25	1.86

Of note, the item in the engagement scale that scored the highest average across modalities is *reaction*, which includes actions such as “liking” and “giving a thumbs up”, while *sharing* has the lowest scores.

Additionally, a positive correlation was found between credibility and engagement, $r = 0.48$ (see **Figure 6**).

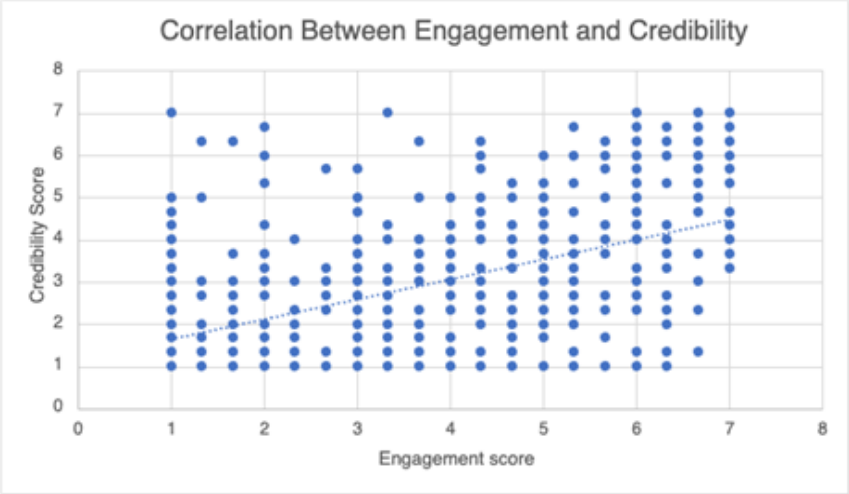


Figure 6. *Correlation Between Engagement and Perceived Credibility*

4.3 The Effect of Warnings on Engagement

H1 predicted that the presence of a warning would negatively affect the intent to engage with the content. To test this hypothesis, an analysis of covariance (ANCOVA) was conducted with engagement as the dependent variable, warning condition (present vs. absent) as the independent variable, and modality (video, audio, text) as a fixed factor. Political orientation, environmental attitude, and science media literacy were included as covariates.

The ANCOVA revealed a significant main effect of warning on engagement, $F(1, 532) = 6.73$, $p = .01$ (see **Table 5**), indicating that participants who were exposed to a warning (*adjusted M* = 2.82, *SE* = 0.10) reported significantly lower willingness to engage with the content compared to those who were not (*adjusted M* = 3.18, *SE* = 0.10). These means are adjusted for political orientation, environmental attitude, and science media literacy. **Thus, H1 is supported.**

In addition, several covariates had significant effects on engagement: environmental attitude, F (1, 532) = 32.94, p < .001, science media literacy, F (1, 532) = 12.46, p <.001, and political orientation, F (1, 532) = 48.26, p <.001. The direction of the effect was positive for all three. Specifically, participants with stronger pro-environmental attitudes (b = 0.435), higher science media literacy (b = 0.267), and more right-leaning political orientation (b = 0.312) reported higher willingness to engage with the content.

Finally, the main effect of modality (p = .06) and its interaction with warning (p = .07) were marginally significant, suggesting that the effect of warnings on engagement may vary across modalities. Further post hoc analyses explore this possibility (see **Section 4.6.1**).

Table 5.

ANCOVA Results for the Direct Effect of Warning on Engagement

Predictor	F (df_1, df_2)	p	η^2
Warning	6.73 (1, 532)	.010	.01
Modality	2.82 (2, 532)	.060	.01
Warning x Modality	2.659 (2, 532)	.071	.01
Environmental attitude	32.94 (1, 532)	< .001	.06
Science media literacy	12.46 (1, 532)	< .001	.02
Political Orientation	48.26 (1, 532)	< .001	.08

4.4 The Indirect Effect of Warning on Engagement Through Credibility

H2 proposed that the effect of a warning on engagement would be mediated by the perceived credibility of the content. Specifically, warnings were expected to reduce perceived credibility, which would, in turn, reduce participants' willingness to engage with the content.

To test the first part of this pathway, an ANCOVA was conducted with perceived credibility as the dependent variable, and warning and modality as fixed factors. Political orientation,

environmental attitude, and science media literacy were included as covariates. The analysis revealed a significant main effect of warning on credibility, F (1, 532) = 18.79, p <.001 (see **Table 6**), indicating that participants who received a warning (*adjusted* M = 3.62, SE = 0.10) rated the content as less credible than those who did not (*adjusted* M = 4.24, SE = 0.11; see **Figure 7**). These adjusted means account for the effects of the covariates.

Table 6.

ANCOVA Results for the Direct Effect of Warning on Credibility

Predictor	F (df_1, df_2)	p	η^2
Warning	18.79 (1, 532)	< .001	.03
Modality	1.60 (2, 532)	.20	.01
Warning x Modality	.07 (2, 532)	.93	.00
Environmental attitude	.292 (1, 532)	.59	.00
Science media literacy	2.31 (1, 532)	.13	.00
Political Affiliation	79.86 (1, 532)	< .001	.13

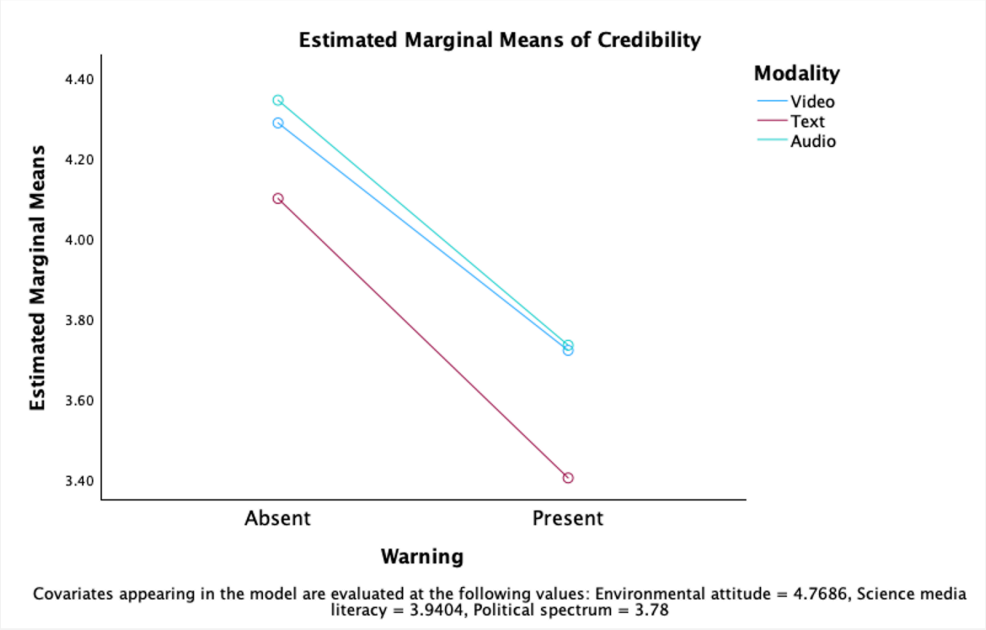


Figure 7. *Estimated Marginal Means of Credibility Across Warning Conditions, by Modality*

Among the control variables, political affiliation had a strong and significant effect on credibility, F (1, 532) = 79.86, p <.001, whereas environmental attitude, F (1, 532) = 0.29, p = .589, and science media literacy, F (1, 532) = 2.31, p =.129, were not significant predictors. More specifically, political affiliation had a positive effect (b = 0.41), signaling that the more right-leaning an individual, the more credible they found the content.

To test the full mediation pathway, we used the PROCESS macro (Hayes, 2017; Model 7; 5,000 resamples). The model examined whether credibility mediated the relationship between warning exposure (X) and engagement (Y), with modality as a moderator on the path from warning to credibility (X → M). Control variables included political orientation, environmental attitude, and science media literacy. Results showed that credibility significantly and positively predicted engagement, b = 0.46, SE = 0.04, p < .001, indicating that higher credibility led to stronger willingness to engage (see **Table 7**).

Table 7.

Summary of Regression Coefficients for Engagement (Model 7)

Predictor	Estimate	*SE*	95% CI		*p*
			LL	*UL*	
Constant	-2.53	0.46	-3.43	-1.62	<.001
Warning	-0.07	0.13	-0.32	0.18	.56
Credibility	0.46	0.04	0.38	0.53	<.001
Environ. Attitude	0.41	0.07	0.28	0.54	<.001
Science Media Lit.	0.34	0.07	0.21	0.47	<.001
Politics	0.13	0.04	0.04	0.21	.003

Note. CI = confidence interval; LL = lower limit; UL = upper limit.

The indirect effect of warning on engagement through credibility was also significant across all three modalities, supporting the mediation hypothesis: video (Effect = -0.26, 95% CI [-0.51,

-0.02]); text (Effect = -0.32, 95% CI [-0.55, -0.10]); and audio (Effect = -0.28, 95% CI [-0.51, -0.06]) (see **Table 8**).

Together, these findings provide **strong support for H2**, demonstrating that the effect of warnings on engagement is mediated by changes in perceived credibility.

Table 8.

Indirect Effects for Each Modality (Model 7)

Modality	Effect	*SE*	95% *CI*	
			LL	*UL*
Video	-0.26	0.12	-0.51	-0.02
Text	-0.32	0.11	-0.55	-0.10
Audio	-0.28	0.12	-0.51	-0.06

Note. CI = confidence interval; LL = lower limit; UL = upper limit.

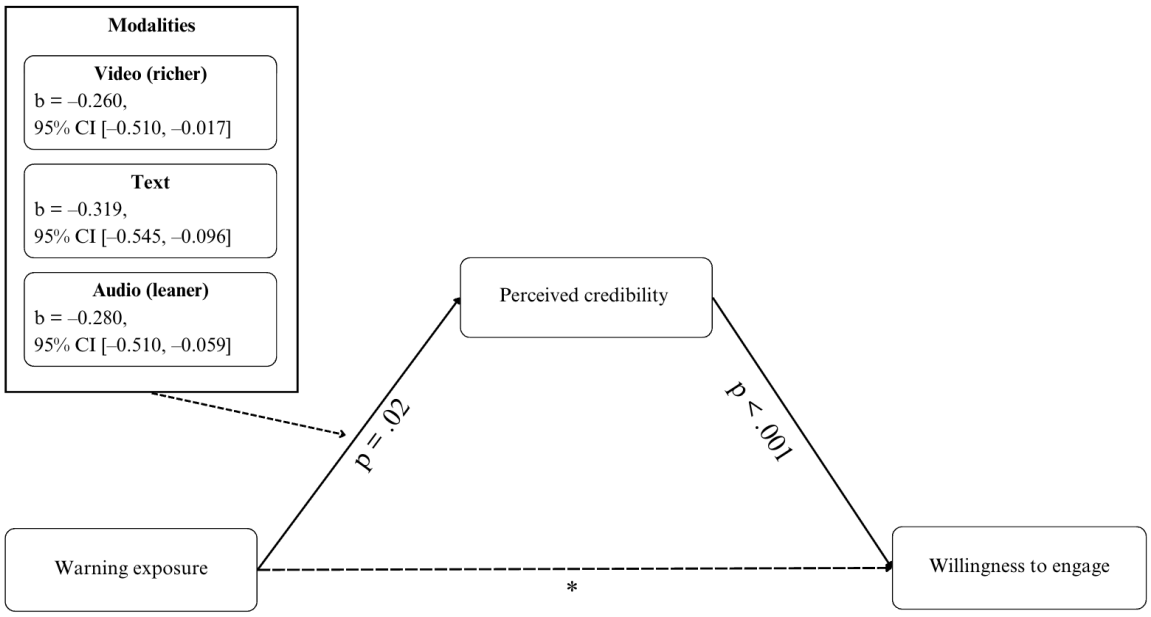
The moderator (modality) was dummy coded with video as the reference category.

4.5 Moderation Effect of Modality

H3 predicted that content modality would moderate the effect of warning exposure on credibility perceptions, with warnings reducing credibility less in richer modalities (e.g., video) than in leaner ones.

To test this hypothesis, a moderated mediation analysis was conducted using the PROCESS macro (Hayes, 2017; Model 7; 5,000 resamples). In the model, warning exposure was entered as the independent variable, perceived credibility as the mediator, engagement as the dependent variable, and modality as the moderator on the path between warning and credibility. Environmental attitude, science media literacy, and political orientation were included as covariates.

As previously shown in **Table 8,** the indirect effect of warnings on engagement through credibility was significant across all three modalities, indicating that credibility consistently mediated the relationship across groups. Accordingly, the index of moderated mediation was not significant when comparing text to the reference category, video (effect = -0.06, 95% CI [-0.38, 0.29]) or audio to the same reference category (effect = -0.02, 95% CI [-0.35, 0.30]), indicating that the strength of the indirect effect did not significantly differ across modalities. In other words, while warnings consistently reduced credibility, this relationship was equally strong across video, text, and audio conditions. **Therefore, H3 is not supported.** These relationships, along with the indirect effect sizes for each modality, are visually summarized in **Figure 8.**



Note. Dotted lines indicate insignificant effects; solid lines indicate significant effect ($p < .05$)

*The direct effect of warning on engagement was significant in an ANCOVA model ($p = .01$), but became non-significant when controlling for perceived credibility (PROCESS Model 7).

Figure 8. Validated Moderated Mediation Model

4.6 Post Hoc Analysis

To further explore the effects of warnings and modality, a series of follow-up ANCOVAs and logistic regressions were conducted. These analyses are exploratory in nature.

4.6.1 Engagement Across Modalities

Post hoc ANCOVAs were conducted to assess whether the effect of warnings on engagement varied across content modalities. The analyses used engagement as dependent variable, and warning exposure (present vs. absent) and modality (video, text, audio) as fixed factors. Political orientation, environmental attitude, and science media literacy were included as covariates.

Results showed a marginally significant main effect of modality on engagement, F (2, 532) = 2.82, p = .06, as well as a marginally significant interaction between modality and warning, F (2, 532) = 2.66, p =.07. Surprisingly, warning exposure did not appear to influence the willingness to engage in the text modality (see **Figure 9**). Post hoc analysis within each modality showed that participants exposed to a warning had slightly higher adjusted engagement scores (M = 2.89, SE = 0.17) than those without warning (M = 2.83, SE = 0.17), although this effect was not statistically significant, F (1,178) = 0.01, p = .92.

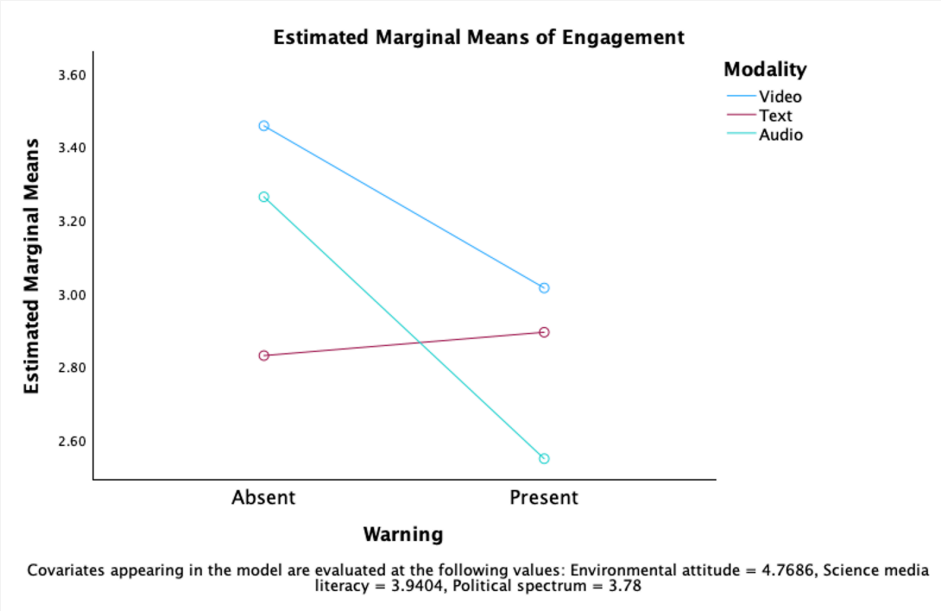


Figure 9. *Estimated Marginal Means of Engagement Across Warning Conditions, by Modality*

In contrast, the video condition exhibited a marginally significant effect of warning on engagement, F (1, 175) = 3.23, p = .07, with adjusted means of M = 3.46 (SE = 0.17) for no warning and M = 3.02 (SE = 0.17) for warning. The audio condition showed a significant effect, F (1, 173) = 9.27, p = .003, with adjusted engagement scores of M = 3.26 (SE = 0.17) without warning and M = 2.55 (SE = 0.17) with warning. In both cases, the presence of a warning reduced the engagement score.

Follow-up ANCOVAs were conducted to compare specific modality pairs and determine whether the effect of warnings on engagement varied between formats. These analyses used a similar model structure as previously described, with engagement as the dependent variable, warning as the independent variable, and political orientation, environmental attitude, and science media literacy as covariates.

The only significant difference in the effect of warning on engagement was found between audio and video, F (1, 351) = 11.43, p <.001, indicating that warnings had a stronger effect in the audio

condition. The difference between audio and text was marginally significant, F (1, 354) = 3.58, p = .06, while the difference between video and text was not statistically significant, F (1, 356) = 1.36, p = .26.

Additional results from the same ANCOVAs revealed that modality only has a significant effect on engagement when comparing text and video, F (1, 356) = 4.74, p = .03. No significant differences were found between text and audio, F (1, 354) = .08, p = .78, and the comparison between audio and video was only marginally significant, F (1, 351) = 3.53, p = .06.

4.6.2 Item-Level Effects on Engagement

To further explore how engagement was affected, post hoc ANCOVAs were conducted on each individual item of the engagement scale, using the same model as before, with warning and modality as fixed factors and the same three covariates included previously.

Results revealed that warning had a significant impact on *sharing*, F (1, 534) = 9.98, p = .002, whereas modality did not, F (2, 534) = 1.67, p = .19. Similarly, warning had a significant effect on *reaction*, F (1, 534) = 7.48, p = .01, while modality only had a marginally significant effect, F (2, 534) = 2.94, p = .05.

In contrast, neither warning, F (1, 534) = 0.66, p = .42, nor modality, F (2, 534) = 2.02, p = .13, significantly affected *commenting*. However, when only video and text were compared, modality showed a marginally significant effect on *commenting*, F (1, 357) = 3.76, p = .05, while warning remained non-significant, F (1, 357) = 0.67, p = .19. These results suggest that the warning condition most strongly influenced the likelihood to share or react, but not to comment.

4.6.3 Credibility Across Modalities

Post hoc ANCOVAs were conducted to compare the effect of warnings on perceived credibility across modality pairs (video/text, video/audio, text/audio). The models included credibility as the

dependent variable, warning exposure as a fixed factor, and political orientation, environmental attitude, and science media literacy as covariates.

Results showed that warnings significantly decreased credibility ratings in all comparisons: video and text, F (1, 356) = 12.83, p <.001, video and audio, F (1, 351) = 10.59, p <.001, and text and audio, F (1, 354) = 14.77, p <.001.

However, the interaction between modality and warning in the overall ANCOVA was not significant, F (2, 532) = 0.07, p = .93 (see **Table 6**), indicating that the effect of warnings on credibility did not vary across modalities. Consistent with this, **Table 9** shows that adjusted credibility scores were lower when a warning was present, regardless of modalities.

Table 9.

Adjusted Means and Standard Errors for Credibility Scores Across Warning Condition, by Modality Pairs

Modality pair	Condition	*Adjusted M*	*SE*
Video/Text	Warning	4.18	0.13
	No Warning	3.54	0.13
Video/Audio	Warning	4.34	0.13
	No warning	3.76	0.13
Text/Audio	Warning	4.22	0.12
	No warning	3.56	0.12

Note. Means are estimated marginal means from ANCOVAs controlling for political orientation, environmental attitude, and science media literacy.

An item-level ANCOVA was conducted on each of the three credibility items with modality and warning as the fixed factors, and political orientation, environmental attitude, and science media literacy as covariates. Results revealed that among the three items of the scale, only *believability*

was significantly influenced by modality, F (2, 532) = 3.71, p = .03. Estimated marginal means showed that participants rated content as most believable in the audio condition (*adjusted M* = 4.32, SE = 0.14), followed by video (*adjusted M* = 4.29, *SE* =0.14), and text (*adjusted M* = 3.84, *SE* = 0.14). However, the main effect of modality on the full credibility scale was not significant overall, F (2, 532) = 1.60, p = .20.

4.6.4 Behavioral Engagement

Finally, a binary logistic regression was conducted to examine the behavioral engagement variable *Want more information*. The dependent variable was participants' response to whether they wanted more information about the content (*yes* = 1, *no* = 0). Predictors included engagement, credibility, warning exposure, and modality (dummy coded with video as the reference category). Political orientation, environmental attitude, and science media literacy were entered as covariates.

Results showed that warning exposure did not significantly predict the likelihood of wanting more information, b = -0.21, *SE* = 0.19, p = .28, *OR* = 0.81, 95% *CI* [0.56, 1.18] (see **Table 10**).

Table 10.

Logistic Regression of the Likelihood of Wanting More Information

Predictor	Estimate	*SE*	*p*	*OR*	95% CI for OR	
					LL	*UL*
Credibility	.20	.07	.002	1.23	1.08	1.40
Modality (1)	.15	.23	.52	1.16	.74	1.83
Modality (2)	.20	.24	.40	1.22	.77	1.94
Warning (1)	-.21	.19	.28	.81	.56	1.18
Engagement	.20	.06	.002	1.22	1.08	1.40
Environ. Attitude	.43	.11	<.001	1.54	1.25	1.91
Science Media Lit.	.42	.11	<.001	1.53	1.23	1.88
Politics	.03	.07	.67	1.03	.90	1.17

Note. OR = odds ratio; CI = confidence interval; LL = lower limit; UL = upper limit.
Modality was dummy coded with video as the reference category. Modality (1) represents text and modality (2) represents audio. Warning was dummy coded with the absent condition (ie., no warning) as the reference category.

In contrast, both higher engagement, $b = 0.20$, $SE = 0.07$, $p = .002$, $OR = 1.22$, 95% *CI* [1.08, 1.39], and higher credibility ratings, $b = 0.20$, $SE = 0.07$, $p = .002$, $OR = 1.23$, 95% *CI* [1.08, 1.39], significantly increased the odds of wanting more information. Specifically, for each one-point increase in engagement or credibility scores, the odds of wanting more information increased by 22.4% and 22.5%, respectively.

Among the control variables, only science media literacy, $b = 0.42$, $SE = 0.11$, $p < .001$, $OR = 1.52$, 95% *CI* [1.23, 1.88], and environmental attitude, $b = 0.43$, $SE = 0.11$, $p < .001$, $OR = 1.54$, 95% *CI* [1.25, 1.91], were significant predictors.

Chapter 5
Discussion

Using an online experimental survey with a between-subjects design, this study investigated the effectiveness of warnings in reducing the perceived credibility of climate-related disinformation and the willingness to engage with it across different content modalities, as part of a moderated mediation model.

Drawing on the Heuristic-Systematic Model (HSM), the MAIN model, and the Media-Richness theory, the research aimed to examine how information processing and heuristic cues interact to shape the perception of and interactions with online disinformation. Specifically, it addressed whether exposure to a warning reduces the willingness to engage with disinformation content (H1), whether credibility mediates this relationship (H2), and whether the format of the content moderates these effects such that warnings would reduce credibility less in richer modalities (e.g., video) than in leaner ones (e.g., text) (H3).

The findings show that warnings significantly reduced engagement overall, supporting H1. Warnings also consistently decreased perceived credibility across all modalities, which in turn lowered engagement intent. This supports the mediation hypothesis (H2) which states that the relationship between warnings and engagement is driven by credibility. However, contrary to what was expected in H3, content modality did not moderate the relationship between warnings and credibility. That is, the effect of warnings on credibility remains stable across video, audio, and text, regardless of the richness of the modality.

Additionally, several control variables influenced measured variables. Environmental attitudes, science media literacy, and political affiliation all positively predicted engagement, while only political affiliation predicted perceived credibility. Specifically, the more right-leaning a participant, the more credible the disinformation was perceived to be and the greater the willingness was to engage with it.

Post hoc analyses provided further nuance to the main findings. While warnings significantly reduced engagement overall, their effect varied across modalities. Specifically, in the text condition, participants who received a warning reported slightly higher (though nonsignificant) engagement scores than those who did not. This contrasted with the audio and video modality, where warnings led to lower engagement. Furthermore, item-level analyses revealed that warnings significantly reduced participants' willingness to share or react to disinformation content, but not to comment on it. Modality also had a significant effect on the credibility scale item *believability,* which was rated higher in the audio and video conditions compared to text. Lastly, both credibility and engagement significantly predicted the behavioral intention to seek more information, reinforcing their role in shaping user behavior.

5.1 Theoretical Implications

This study offers several theoretical implications that supplement and contradict certain elements of the framework this research is built on.

5.1.1 Effectiveness of Warnings and HSM

To start, the results support the effectiveness of warnings in reducing both perceived credibility and, indirectly, engagement with disinformation. This aligns with the HSM, which posits that accuracy motivation encourages systematic processing, leading individuals to scrutinize information more critically and become more resistant to misleading content. The warning used in this study may have served as an external cue that activated this type of motivation. This interpretation is supported by prior research, such as Pennycook, Epstein, et al. (2021), who found that prompting people to consider accuracy increases the quality of the content they share. The authors suggest this is because people are often distracted from considering the accuracy of content, implying that low accuracy motivation may explain the susceptibility to sharing disinformation.

Additionally, the warning may have increased the discrepancy between participants' actual and desired confidence in their ability to assess the content, which, according to the sufficiency

principle, would push people towards more effortful, systematic processing. The increase might have happened because the generic warning prompted participants to pause and reflect on the credibility of the information without offering a definitive judgment on its accuracy, thus encouraging them to reassess their confidence.

Relatedly, this research found evidence that credibility positively influences engagement: the more credible participants perceived the content to be, the more willing they were to engage with it. This is somewhat unexpected given that prior studies suggest people may share content even when they know it is inaccurate (e.g., Chadwick et al., 2018; Molina et al., 2023; Pennycook, Bear, et al., 2020). One possible explanation is that, in a hypothetical scenario like the one used in this study, participants may feel more comfortable indicating that they would not engage with content they rated as non credible than in reality. Another possibility is that, in the absence of the cognitive overload and distractions typically present on social media platforms, participants in this study were more likely to engage in systematic processing, potentially facilitated by the accuracy motivation embedded in the design of the task. Future research should aim to replicate this finding by measuring observed behavior in more naturalistic settings instead.

Unexpectedly, the text modality showed a reversed trend, with slightly higher engagement when a warning was present, although this effect was not statistically significant. While this may point to modality specific differences in how users process warnings in terms of engagement, it may also reflect random variation rather than a meaningful effect. Importantly, credibility decreased fairly consistently across all modalities, indicating that the pattern in the text condition is specific to engagement. This finding aligns with research by Hameleers et al. (2020), which demonstrated that corrective information effectively debunks disinformation in both textual and visual modalities resulting in reduced perceived credibility. It also raises important questions about the influence of modality on engagement behaviors, which should be further investigated through studies that focus on actual user behaviors rather than self-reported intentions.

5.1.2 Absence of Modality Effects

This study did not find evidence to support the hypothesis, aligned with the MAIN model and the Media Richness theory, that content modality would moderate the effect of warnings on credibility. There was also no significant main effect of modality on credibility, reinforcing the unexpected nature of this finding. These results suggest that warnings operate similarly across content formats, regardless of the richness of the medium, challenging the assumption that richer formats are affected differently than leaner ones.

According to the MAIN model's realism heuristic, sensory-rich media, such as video and to a lesser extent audio, should be perceived as more credible because they more closely simulate real life. This view aligns with the Media Richness Theory (Daft & Lengel, 1986), which posits that richer media, offering numerous cues and immediate feedback, are better suited for communicating ambiguous information and may therefore appear more credible in the case of disinformation. In addition, visual information tends to prompt greater processing fluency, making the content easier to assimilate and, consequently, appear more truthful (Vaccari & Chadwick, 2020). As Lang's (2000) Limited Capacity Model of Mediated Message Processing (LC4MP) suggests, richer modalities can also demand more cognitive effort to encode because they engage multiple sensory and cognitive channels simultaneously. This can lead to greater credibility perceptions through an increased reliance on heuristic processing.

Other heuristics, like the old-media heuristic, may counterbalance the realism heuristic by favoring text-based content, which resembles traditional and credible sources of knowledge like newspapers and books. While relevant heuristics may compete, they all imply that modality should have moderated the effect of warnings on credibility one way or another. Yet, despite these theoretical expectations, the present study found no significant differences across modalities, suggesting that modality alone may not consistently shape credibility judgments in the context of warnings.

Still, this finding is consistent with recent work by Hameleers et al. (2022), who found no evidence that deepfakes have stronger persuasive power than textual disinformation. In fact, the

authors observed that deepfakes were sometimes perceived as less credible than text-based disinformation. This aligns with Vaccari & Chadwick (2020) argument that the ubiquity of both fake news and fake news accusations contributes to an environment of uncertainty, ultimately reducing trust in all forms of news on social media.

In the current study, average credibility ratings across all modalities hovered around the midpoint of the scale or lower, pointing to this same climate of skepticism. Hameleers (2024) observed a similar pattern: both deepfakes and cheapfakes (i.e., real videos where the audio has been altered) were rated similarly low on credibility, while text-based disinformation was rated as equally credible as authentic information. The author similarly concludes that this may reflect a broader public uncertainty that leads individuals to assess all information as only moderately credible. Another possible explanation for the ratings in this study is that the disinformation content selected may have simply seemed rather implausible to most participants, limiting credibility regardless of format.

Additionally, a potential limitation that could also explain the absence of modality effects lies in the design of the stimuli. Only the video condition was pretested for clarity and professionalism to select the specific clip used in the main study. The audio and text conditions were directly derived from the video (via audio extraction and transcript) and were not pretested separately.

Notably, the text condition was displayed using basic Qualtrics formatting, which may have lacked the visual credibility cues typically associated with published articles. This disparity may have contributed to lower credibility ratings in the text condition. Audio, which lacked visual context, also relied on text to convey the source of the content. These inconsistencies illustrate a common challenge in multimodal research: achieving consistent levels of realism across modalities and maintaining equivalent information richness.

Because the video presented the content in its intended form, it may have appeared more professionally produced, skewing the ratings. Likewise, the audio condition, though it lacked visual cues, retained the original voice and tone, which may have appeared more professional as

well. These factors may have inadvertently advantaged the video and audio conditions compared to text.

At the same time, the text condition, by requiring more effortful reading and offering fewer visual cues, may have encouraged more systematic processing. This, alongside the barebone formatting, could explain the trend toward lower (albeit non-significantly so) credibility ratings in the text condition.

Interestingly, while overall credibility ratings were not significantly affected by modality, *believability*, one of the three components of the credibility scale, was. Even though the scale was found to be reliable, this pattern suggests that believability may tap into a slightly different cognitive process. Unlike accuracy and authenticity, which may prompt more objective evaluations (e.g., "Is this factually correct?" and "Does this seem genuine?"), believability may reflect a more internal or hypothetical judgment (e.g., "Would I or others find this believable?"). This interpretation seems to be supported by Appelman & Sundar (2016), who note that while all three dimensions are subjective as they are self-reported, believability may be considered even more subjective. As such, this dimension may be particularly susceptible to heuristic cues triggered by content modality, such as the realism heuristic.

In conclusion, these findings challenge a key assumption of the MAIN model, which is that modality shapes credibility perceptions through heuristic cues. The lack of significant differences across modalities indicates that such cues may not be very influential in the context of disinformation warnings, particularly when overall credibility is already low. This could reflect today's information landscape, where individuals are skeptical of most online content, whether true or false. The fact that only believability was influenced by modality further suggests that the various aspects of credibility may be differently affected by content format.

These results also highlight a theoretical gap in the HSM, which does not explicitly account for the role of modality in influencing information processing. While the HSM outlines how heuristic and systematic processing affect credibility judgment, it overlooks how different media

formats may engage these routes differently. Integrating the HSM framework with Lang's (2000) LC4MP could offer a more nuanced understanding of how modality interacts with cognitive efforts, heuristic activation, and credibility evaluation. Such a refinement would provide a stronger framework for examining the complex ways people engage with multimodal disinformation.

5.1.3 Political Affiliation and Disinformation

Lastly, although political affiliation was not part of this study's experimental manipulations, the results align with prior research indicating that individuals who identify as more right-leaning were likely to perceive disinformation as more credible and to express greater willingness to engage with it than left-leaning individuals (Lewandowsky, 2021; Renault et al., 2025; Tam & Chan, 2023). Given the focus of this thesis on climate change, a topic that tends to be more politically polarizing, these findings are unsurprising.

However, the underlying mechanisms behind this asymmetry remain unclear. One possibility is that a general distrust in mainstream media may drive individuals to oppose dominant narratives, especially on complex issues like climate change that are difficult to verify independently (Lewandowsky, 2021). Additionally, the psychological discomfort caused by large-scale threats may lead some individuals to adopt conspiracy theories as a means of coping with uncertainty (Booth et al., 2024). The role of influential conservative figures, such as President Donald Trump, in promoting or legitimizing such conspiracy theories may further reinforce these beliefs.

From the perspective of the HSM, these findings seem to indicate that right-leaning individuals are less inclined to engage in systematic processing when presented with disinformation that aligns with their beliefs. This may be because their processing is guided by defense motivation (aimed at protecting one's identity) or impression motivation (aimed at presenting a socially favorable image), rather than by accuracy motivation. According to the sufficiency principle, when trusted political figures reinforce these views, the gap between individuals' actual and desired confidence may shrink, reducing the motivation to process information more thoroughly.

This reinforcement may inflate their actual confidence, giving them the impression that they understand the issue sufficiently.

These findings highlight the importance of accounting for political orientation in disinformation research, at least as a control variable. While this asymmetry warrants further exploration, it is noteworthy that warnings remained effective overall in reducing both perceived credibility and engagement. This suggests that, despite ideological differences, warnings may retain effectiveness across the political spectrum, though future studies should test this moderation effect directly.

Environmental attitudes and science media literacy were also significant positive predictors of the willingness to engage but not of perceived credibility. In other words, participants with stronger pro-environmental attitudes and higher media literacy scores were more likely to report a willingness to engage with the disinformation content. One possible explanation is that these individuals are more likely to engage with the content in order to question, clarify, or refute it. This interpretation is reinforced by this study's binary logistic regression results, which revealed that participants with higher scores on these scales were more likely to indicate that they "Want more information", potentially as a way to refine or challenge their understanding of the content. Given that both scales were self-reported, some level of bias is also possible.

5.2 Practical Implications

Beyond its theoretical contributions, this research addresses an important real-world issue: the very real threat of disinformation. As social media platforms remain central to its dissemination, practical strategies to mitigate its impact are increasingly important for both policymakers and social media companies.

This study sheds light on the efficacy and limitations of warnings, a widely used intervention, offering insights into their effectiveness across modalities and in relation to different engagement

behaviors. From a managerial perspective, several actionable insights emerge from these findings.

5.2.1 General Effectiveness of Warnings and Content Moderation Strategies

A key finding of this research is that warnings do not need to be tailored to specific content to be effective, since even generic warnings significantly reduced both perceived credibility and engagement intent. Prior research has found that warnings are more effective when they include content correcting the misleading information and explain the deceptive techniques used (Lewandowsky & van der Linden, 2021). The present findings extend this work by showing that even generic warnings can meaningfully reduce engagement and credibility. This is consistent with Jalbert et al. (2020), who found that even basic reminders that not all information is true, similar to the warning used in this study, can reduce belief in false content. According to the authors, such interventions may help mitigate the delay between the initial spread of disinformation and its fact-checking.

This finding has important practical relevance as social media platforms reevaluate their content moderation strategies. For instance, Meta recently announced that they would phase out third-party fact-checkers in favor of a community-based system similar to the "Community Notes" used on Twitter (Kaplan, 2025). Under this system, users flag questionable content and write notes explaining the reason why they did. Notes are then voted on for helpfulness by other users. Meta CEO Mark Zuckerberg stated that this shift was partly in response to concerns that third-party fact-checking organizations may be ideologically biased, particularly against conservative content (Zahn, 2025).

While this shift was framed as a way to be more neutral, a recent study suggests that even crowdsourcing systems tend to flag Republican content more than Democrat content, raising the question of whether bias stems from the moderation method or the content itself (Renault et al., 2025). Another study has found that while individuals are indeed polarized when rating news sources on social media, the effect appears symmetrical across the political spectrum, effectively

offsetting bias when aggregated (Epstein et al., 2020). This finding supports the explanation that disparities may be driven by the nature of the content itself rather than by the content moderation strategy.

Another potential challenge with the Community Notes model is that not all content gets flagged, which may trigger the implied truth effect, where unlabeled content is perceived as accurate simply because it lacks a warning (Pennycook, Bear, et al., 2020). To mitigate this risk, social media platforms should aim for a consistent application of warning labels. As suggested by Jalbert et al. (2020) and supported by the present findings, a general warning could be used preemptively and later combined with crowdsourced notes once the content has been reviewed.

5.2.2 Modality-Agnostic Nature of Warnings

Another important takeaway is that warnings seem to be modality-agnostic. In other words, the present findings suggest that content format, whether video, audio, or text, does not significantly moderate the relationship between warnings and perceived credibility. It indicates that the effectiveness of warnings is stable across different formats of disinformation.

This result aligns with Hameleers et al. (2020), who similarly reported that the modality of a fact-checking message did not impact its effectiveness. The authors explained that text-only messages may be particularly effective because fact-checkers use concise, evidence-based argumentations, which facilitates comprehension. From a practical standpoint, these findings suggest that even simple, text-based warnings can be deployed effectively across a variety of content types and platforms. This has important implications for businesses as text-based warnings are easier to rapidly implement at scale, are more cost-efficient to produce, and are adaptable to a wide range of digital environments.

5.2.3 Warning Design

This study also contributes to a better understanding of what constitutes disruption in the context of warnings. As discussed in the literature review, warnings tend to be more effective when they

interrupt user interaction with content, such as through modal windows that must be closed to proceed (Mende et al., 2024). In this case, participants were shown a general warning embedded on the instruction page and were required to click on a button to continue. While relatively moderate in intrusiveness, as confirmed by pretest results showing that around 15% of participants did not notice the warning, this level of disruption was still sufficient to significantly reduce perceived credibility and, indirectly, engagement intentions. This demonstrates that even moderate forms of disruptions can be impactful, offering a good balance between user-friendliness and intervention.

The findings of this research highlight that even simple text-only interventions, when moderately disruptive and consistently applied, can be effective at mitigating the impact of disinformation. When combined with more sophisticated solutions, such as Community Notes or third-party fact-checking, they have the potential to prevent further biases.

5.2.4 Engagement Motivation and Behavioral Differences

Finally, post hoc analyses examined each item in the engagement scales to identify potential patterns, as engagement behaviors may be driven by different motivations. The results showed that warnings significantly reduced willingness to share and react to the content, but had no statistical effect on willingness to comment.

One possible explanation is that commenting can serve corrective functions, such as mocking or disputing false information. People may therefore still want to comment on flagged content to provide additional context or clarification. In contrast, reacting or sharing may be more easily interpreted as endorsement, making individuals more hesitant to engage in these behaviors when a warning is present. This interpretation is supported by Molina et al. (2023), who found that engagement behavior differed depending on whether the content was true or false: participants were more likely to react to real content, whereas they were more likely to comment on false content due to feeling of unease toward it.

These possible differences in engagement motivation are important to consider as they suggest that future research should distinguish between forms of engagement when designing measures for their studies. They also highlight that interventions should be tailored to specific engagement behaviors, acknowledging that engagement is a nuanced behavior that is not inherently positive or negative.

5.3 Limitations and Future Research

This research has multiple limitations. First, the study was conducted online in a controlled setting, which may limit its ecological validity. Participants' interactions with the content likely did not fully replicate how they encounter and process information during everyday social media use. For example, typical social media distractions such as ads, visible user engagement (e.g., comments, likes, etc.), and more importantly, viewing multiple posts simultaneously were absent. As a result, participants may have evaluated their engagement intent and credibility perceptions with more deliberation than they would in a natural setting. Future research could replicate this study within real or simulated social media environments to better reflect user behavior.

Second, the sample may reflect biases related to online participation. Although Prolific has been shown to yield higher quality data than other online recruitment platforms (Peer et al., 2022), it remains difficult to verify who is responding and how attentive they are past the attention checks. Lab-based research could help address these concerns by providing greater control over participant identification and engagement with the study. Additionally, the sample used in the study was not selected to ensure statistical representativeness of the targeted populations, which limits the generalizability of the findings.

Third, the study did not account for the type of device participants used. This means that participants may have completed the study using desktop or mobile devices, which introduces a possible confounding variable. While the impact of device type on information processing

remains debated (Sundar et al., 2025), some research suggests that mobile users may be more susceptible to disinformation due to reduced processing time, which can lead to decreased attention to content accuracy (Liao et al., 2023). Future research should examine whether the effectiveness of warnings and the impact of modality differ by device type.

Fourth, most of the measures used in this study were self-reported, a method known to be prone to biases (Podsakoff et al., 2003). For example, participants may be influenced by the social desirability bias, leading them to provide answers they believe are more socially acceptable rather than those that reflect their true attitudes or intentions. Additionally, people are often poor at predicting their own future behavior, especially in hypothetical scenarios (Poon et al., 2014). Future research may benefit from using behavioral measures of engagement, such as actual sharing behavior or click data, to more accurately assess the effects of disinformation and warnings.

Finally, this study opens up several promising research avenues. As the present findings rely on self-reported measures, they offer insights into what participants perceive or intend to do, but not necessarily why that is the case. To address this gap, qualitative methods could be used to explore participants' motivation for engagement and interpretation of credibility cues. Additionally, physiological measures, such as eye tracking or heart rate variability could help understand the cognitive and emotional mechanism underlying user responses. Future studies could also test different types of warnings, varying in style, specificity, placement, or level of disruption, to further clarify the boundaries of what is or is not effective.

Chapter 6
Conclusion

This study investigated the influence of warnings on the willingness to engage with disinformation content and its perceived credibility. It also examined whether content modality moderated the effects of warnings on disinformation credibility. To test this, a between-subjects online survey experiment was conducted in which participants were exposed to disinformation content in one of three formats (video, audio, text), half of them with a warning about the content. Participants were then asked to rate both the content's credibility and their willingness to engage with it.

The findings showed that warnings effectively reduced perceived credibility, and, indirectly, willingness to engage with the content. However, there was no statistical evidence that modality significantly influenced the effect of warnings on credibility, suggesting that the intervention worked similarly across content formats.

6.1 Theoretical Contributions

From a theoretical standpoint, the results align with Pennycook, Epstein, et al. (2021), who found that prompting people to focus on content accuracy reduces the likelihood of sharing disinformation. This supports the idea that warnings may activate the accuracy motivation, as defined in the Heuristic-Systematic Model (HSM), which drives individuals to process information with the goal of making well-informed judgments. Additionally, the presence of a warning may increase individuals' desired confidence, which, according to the sufficiency principle in the HSM, further encourages systematic processing.

Credibility was also found to mediate the relationship between warnings and the willingness to engage with content. This suggests that credibility is not only an outcome of information processing, but also a key driver of behavioral intentions on social media. By connecting

credibility with engagement behaviors like commenting, sharing, and reacting, this study extends the application of the HSM into the domain of user interaction with persuasive content.

At the same time, the findings challenge assumptions in both the MAIN model and the Media Richness Theory. The MAIN model posits that modality shapes credibility through heuristic cues, such as the realism heuristic, through which content that appears more realistic is perceived as more credible. Media Richness Theory predicts that richer media (such as video) are more effective at conveying complex or ambiguous information because they provide multiple cues and immediate feedback, which can enhance credibility. Contrary to these expectations, no significant moderation effect or main effects of modality were found. This suggests that the heuristic cues triggered by different modalities (audio, text, video) may not differ meaningfully in their impact on credibility, despite different levels of media richness. However, the finding that *believability* alone, one the three items of the credibility scale, was significantly influenced by modality suggests that this dimension of credibility may be more sensitive to modality-specific heuristic cues.

Finally, this study contributes to the growing body of evidence that political affiliation significantly influences credibility judgments, particularly in the context of polarized topics such as climate change. From the perspective of the HSM, this may reflect the role of defense motivation, where individuals process information in a way that protects their ideological identity, or impression motivation, where processing is done to present a socially favorable image, rather than pursuing accuracy. This may help explain why right-leaning participants were more likely to perceive climate-related disinformation as credible and expressed greater willingness to engage with it.

6.2 Practical Contributions

In terms of practical implications, this study offers valuable insights for social media and digital platforms seeking to mitigate the spread of disinformation. The findings suggest that even a generic, text-only warning, like the one used in this study, can effectively reduce the perceived

credibility of disinformation and users' willingness to engage with it. Since this type of intervention works across modalities, it is easily scalable and can complement more targeted strategies already in place. Importantly, a broad application of warnings may help counter the implied truth effect, in which unflagged disinformation content tends to be perceived as accurate because its lack of a warning is interpreted as a credibility cue (Pennycook, Bear, et al., 2020).

While this study did not manipulate the level of warning intrusiveness directly, the findings suggest that even moderately disruptive warnings, such as requiring a click-through from an instruction page, can be effective. Despite its relatively low intrusiveness, it still significantly reduced credibility and engagement. This supports previous findings that interruptions in user flow can enhance the effectiveness of interventions against disinformation (Mende et al., 2024).

Additionally, this research highlights the importance of distinguishing different forms of engagement behavior, rather than assuming they all reflect the same intentions. The findings showed that warnings did not have a statistically significant effect on the action of commenting, while it did on the two other items of the scale, reacting and sharing. Since the motivation behind these actions likely vary depending on the context and individual, organizations should avoid treating all engagement metrics as equivalent. Recognizing this nuance is essential for designing more effective interventions and for interpreting engagement data more meaningfully.

Overall, this study underscores the need for a consistent application of credibility cues across content to avoid the unintended reinforcement of disinformation. It also confirms that reducing the credibility of content has the potential to influence behavior, such as reducing the willingness to engage with disinformation online.

6.3 Limitations and Future Research

While this study provides meaningful insights into the effects of warnings and modality on disinformation credibility and engagement, several limitations should be acknowledged. These also highlight important avenues for future research.

A key limitation is that the study was done in an artificial context. Participants were exposed to disinformation content in a controlled survey environment, free from the competing stimuli and distractions typical of social media platforms. Replicating the study in a more ecologically valid setting would help assess how these findings translate to real-world conditions.

Second, only the video stimulus was pretested for visual professionalism. The audio and text versions were adapted from the video but were not evaluated for visual quality or for their presentation on Qualtrics. This may have affected how credible or engaging participants found each modality. Future research should ensure that all modalities are pretested to achieve comparable levels of perceived professionalism. Alternatively, studies could start with a different modality (e.g., text) and derive the other formats from it to verify whether the observed patterns hold.

Another methodological limitation is the reliance on self-reported measures of credibility and engagement, which may not accurately reflect actual behavior. Incorporating behavioral measures, such as real engagement and physical attention to the content, would provide more solid proofs of engagement. Additionally, the device on which participants completed the study (e.g., mobile versus desktop) was not recorded, even though it may influence how warnings and modalities are processed. Future studies should consider device type as a relevant variable to examine whether perceived credibility and willingness to engage vary depending on screen size and mode of interaction.

Further studies could also explore different warning types, varying both in content and design, to determine which configurations are most effective. Finally, it would be valuable to test this intervention across a wider range of topics with varying levels of political polarization, to evaluate whether its effectiveness generalizes beyond the context of climate change.

In conclusion, this thesis contributes to the literature on disinformation by demonstrating that even generic warning labels can effectively reduce how credible disinformation is perceived, as

well as individuals' willingness to engage with it on social media. Contrary to expectations, this study also found that the effect of warnings on credibility did not vary meaningfully across content modalities.

Disinformation is unlikely to disappear, especially as social media continues to facilitate its production and spread. As such, it is critical to develop scalable interventions that can be applied in a wide range of contexts. This research highlights that even simple, one-size-fits-all text-based warnings can be effective tools for platforms seeking to fight back against disinformation.

Bibliography

Aïmeur, E., Amri, S., & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: A review. *Social Network Analysis and Mining*, *13*(1), 30. https://doi.org/10.1007/s13278-023-01028-5

Ali, K., Li, C., Zain-ul-abdin, K., & Zaffar, M. A. (2021). Fake news on Facebook: Examining the impact of heuristic cues on perceived credibility and sharing intention. *Internet Research*, *32*(1), 379–397. https://doi.org/10.1108/INTR-10-2019-0442

American Psychological Association. (2024). *Potential risks of content, features, and functions: The science of how social media affects youth* (p. 6). https://www.apa.org/topics/social-media-internet/youth-social-media-2024

Amnesty International. (2022, September 29). *Myanmar: Facebook's systems promoted violence against Rohingya; Meta owes reparations – new report*. Amnesty International. https://www.amnesty.org/en/latest/news/2022/09/myanmar-facebooks-systems-promoted-violence-against-rohingya-meta-owes-reparations-new-report/

Angelucci, C., & Prat, A. (2024). Is Journalistic Truth Dead? Measuring How Informed Voters Are about Political News. *American Economic Review*, *114*(4), 887–925. https://doi.org/10.1257/aer.20211003

Appelman, A., & Sundar, S. S. (2016). Measuring Message Credibility: Construction and Validation of an Exclusive Scale. *Journalism & Mass Communication Quarterly*, *93*(1), 59–79. https://doi.org/10.1177/1077699015606057

Austin, E. W., Austin, B. W., Bolls, P. D., Edwards, Z. M., Domgaard, S. K., Mu, D., O'Donnel, N. H., Payne, C., Rose, P., & Sheftel, A. (2023). *Getting to the Heart and Mind of the*

Matter: A Toolkit to Build Confidence as a Trusted Messenger of Health Information—Washington State University (2nd edition). University of Washington. https://rex.libraries.wsu.edu/esploro/outputs/other/Getting-to-the-Heart-and-Mind/99901131434601842

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://doi.org/10.1037/xge0000729

Booth, E., Lee, J., Rizoiu, M.-A., & Farid, H. (2024). Conspiracy, misinformation, radicalisation: Understanding the online pathway to indoctrination and opportunities for intervention. *Journal of Sociology*, *60*(2), 440–457. https://doi.org/10.1177/14407833241231756

Chadwick, A., Vaccari, C., & O'Loughlin, B. (2018). Do tabloids poison the well of social media? Explaining democratically dysfunctional news sharing. *New Media & Society*, *20*(11), 4255–4274. https://doi.org/10.1177/1461444818769689

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, *39*(5), 752–766. https://doi.org/10.1037/0022-3514.39.5.752

Chaiken, S., & Ledgerwood, A. (2012). A theory of heuristic and systematic information processing. In *Handbook of theories of social psychology, Vol. 1* (pp. 246–266). Sage Publications Ltd. https://doi.org/10.4135/9781446249215.n13

Colliander, J. (2019). "This is fake news": Investigating the role of conformity to other users' views when commenting on and spreading disinformation in social media. *Computers in*

Human Behavior, *97*, 202–215. https://doi.org/10.1016/j.chb.2019.03.032

Daft, R. L., & Lengel, R. H. (1986). Organizational Information Requirements, Media Richness and Structural Design. *Management Science*, *32*(5), 554–571. https://doi.org/10.1287/mnsc.32.5.554

DeSmog Climate Disinformation Database. (n.d.-a). CO2 Coalition. *DeSmog*. Retrieved March 31, 2025, from https://www.desmog.com/co2-coalition/

DeSmog Climate Disinformation Database. (n.d.-b). Organization. *DeSmog*. Retrieved April 28, 2025, from https://www.desmog.com/organization/

Di Domenico, G. D., Nunan, D., & Pitardi, V. (2022). Marketplaces of Misinformation: A Study of How Vaccine Misinformation Is Legitimized on Social Media. *Journal of Public Policy & Marketing*. https://doi.org/10.1177/07439156221103860

Di Domenico, G. D., Sit, J., Ishizaka, A., & Nunan, D. (2021). Fake news, social media and marketing: A systematic review. *Journal of Business Research*, *124*, 329–341. https://doi.org/10.1016/j.jbusres.2020.11.037

Diaz Ruiz, C. (2023). Disinformation on digital media platforms: A market-shaping approach. *New Media & Society*, 14614448231207644. https://doi.org/10.1177/14614448231207644

Diaz Ruiz, C., & Nilsson, T. (2023). Disinformation and Echo Chambers: How Disinformation Circulates on Social Media Through Identity-Driven Controversies. *Journal of Public Policy & Marketing*, *42*(1), 18–35. https://doi.org/10.1177/07439156221103852

Ecker, U. K. H., Lewandowsky, S., & Chadwick, M. (2020). Can corrections spread misinformation to new audiences? Testing for the elusive familiarity backfire effect.

COGNITIVE RESEARCH-PRINCIPLES AND IMPLICATIONS, *5*(1), 41. https://doi.org/10.1186/s41235-020-00241-6

Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & 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

Epstein, Z., Pennycook, G., & Rand, D. (2020). Will the Crowd Game the Algorithm? Using Layperson Judgments to Combat Misinformation on Social Media by Downranking Distrusted Sources. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–11. https://doi.org/10.1145/3313831.3376232

Fisher, J. T., Hopp, F. R., & Weber, R. (2019). *Modality-Specific Effects of Perceptual Load in Multimedia Processing | Article | Media and Communication*. https://www.cogitatiopress.com/mediaandcommunication/article/view/2388

Gottfried, J. L. and J. (2022, October 27). U.S. adults under 30 now trust information from social media almost as much as from national news outlets. *Pew Research Center*. https://www.pewresearch.org/short-reads/2022/10/27/u-s-adults-under-30-now-trust-information-from-social-media-almost-as-much-as-from-national-news-outlets/

Hameleers, M. (2024). Cheap Versus Deep Manipulation: The Effects of Cheapfakes Versus Deepfakes in a Political Setting. *International Journal of Public Opinion Research*, *36*(1), edae004. https://doi.org/10.1093/ijpor/edae004

Hameleers, M., & Meer, T. G. L. A. V. der. (2021). The Scientists Have Betrayed Us! The Effects of Anti-Science Communication on Negative Perceptions Toward the Scientific

Community. *International Journal of Communication*, *15*(0), Article 0.

Hameleers, M., Powell, T. E., Van Der Meer, T. G. L. A., & Bos, L. (2020). A Picture Paints a Thousand Lies? The Effects and Mechanisms of Multimodal Disinformation and Rebuttals Disseminated via Social Media. *Political Communication*, *37*(2), 281–301. https://doi.org/10.1080/10584609.2019.1674979

Hameleers, M., van der Meer, T. G. L. A., & Dobber, T. (2022). You Won't Believe What They Just Said! The Effects of Political Deepfakes Embedded as Vox Populi on Social Media. *Social Media + Society*, *8*(3), 20563051221116346. https://doi.org/10.1177/20563051221116346

Hayes, A. F. (2017). *Introduction to Mediation, Moderation, and Conditional Process Analysis*. Guilford Publications. http://afhayes.com/introduction-to-mediation-moderation-and-conditional-process-analysis.html

Imhoff, R., & Lamberty, P. (2020). A Bioweapon or a Hoax? The Link Between Distinct Conspiracy Beliefs About the Coronavirus Disease (COVID-19) Outbreak and Pandemic Behavior. *Social Psychological and Personality Science*, *11*(8), 1110–1118. https://doi.org/10.1177/1948550620934692

Ipsos. (2023). Global views on A.I. and disinformation: Perception of Disinformation Risks in the Age of Generative A.I. *New Zealand*.

Islam, M. S., Kamal, A.-H. M., Kabir, A., Southern, D. L., Khan, S. H., Hasan, S. M. M., Sarkar, T., Sharmin, S., Das, S., Roy, T., Harun, M. G. D., Chughtai, A. A., Homaira, N., & Seale, H. (2021). COVID-19 vaccine rumors and conspiracy theories: The need for

cognitive inoculation against misinformation to improve vaccine adherence. *PLoS ONE*, *16*(5), e0251605. https://doi.org/10.1371/journal.pone.0251605

Jalbert, M., Newman, E., & Schwarz, N. (2020). Only Half of What I'll Tell You is True: Expecting to Encounter Falsehoods Reduces Illusory Truth. *Journal of Applied Research in Memory and Cognition*, *9*(4), 602–613. https://doi.org/10.1016/j.jarmac.2020.08.010

Jones-Jang, S. M., Mortensen, T., & Liu, J. (2021). Does Media Literacy Help Identification of Fake News? Information Literacy Helps, but Other Literacies Don't. *American Behavioral Scientist*, *65*(2), 371–388. https://doi.org/10.1177/0002764219869406

Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making*, *8*(4), 407–424. https://doi.org/10.1017/S1930297500005271

Kaplan, J. (2025, January 7). Meta—More Speech and Fewer Mistakes. *Meta Newsroom*. https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/

Kaufman, M. (2020, July 13). *The Carbon Footprint Sham*. Mashable. https://mashable.com/feature/carbon-footprint-pr-campaign-sham

Lang, A. (2006). Using the Limited Capacity Model of Motivated Mediated Message Processing to Design Effective Cancer Communication Messages. *Journal of Communication*, *56*(s1), S57–S80. https://doi.org/10.1111/j.1460-2466.2006.00283.x

Lee, J., & Shin, S. Y. (2022). Something that They Never Said: Multimodal Disinformation and Source Vividness in Understanding the Power of AI-Enabled Deepfake News. *Media Psychology*, *25*(4), 531–546. https://doi.org/10.1080/15213269.2021.2007489

Lewandowsky, S. (2021). Climate Change Disinformation and How to Combat It. *Annual Review of Public Health*, *42*, 1–21.

https://doi.org/10.1146/annurev-publhealth-090419-102409

Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition*, *6*(4), 353–369. https://doi.org/10.1016/j.jarmac.2017.07.008

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

Liao, M., Wang, J., Chen, C., & Sundar, S. S. (2023). Less vigilant in the mobile era? A comparison of information processing on mobile phones and personal computers. *New Media & Society*, 14614448231209475. https://doi.org/10.1177/14614448231209475

Luo, M., Hancock, J. T., & Markowitz, D. M. (2022). Credibility Perceptions and Detection Accuracy of Fake News Headlines on Social Media: Effects of Truth-Bias and Endorsement Cues. *Communication Research*, *49*(2), 171–195. https://doi.org/10.1177/0093650220921321

Martel, C., Pennycook, G., & Rand, D. G. (2020). Reliance on emotion promotes belief in fake news. *Cognitive Research: Principles and Implications*, *5*(1), 47. https://doi.org/10.1186/s41235-020-00252-3

Martel, C., & Rand, D. G. (2023). Misinformation warning labels are widely effective: A review of warning effects and their moderating features. *Current Opinion in Psychology*, *54*, 101710. https://doi.org/10.1016/j.copsyc.2023.101710

Mende, M., Ubal, V. O., Cozac, M., Vallen, B., & Berry, C. (2024). Fighting Infodemics: Labels as Antidotes to Mis- and Disinformation?! *Journal of Public Policy & Marketing*, *43*(1),

31–52. https://doi.org/10.1177/07439156231184816

Meta. (2022, October 24). Learn More About the Types of Techniques Used to Misrepresent Climate Science Online. *Meta Sustainability*. https://sustainability.atmeta.com/blog/2022/10/24/climate-science-literacy-initiative/

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

Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and Heuristic Approaches to Credibility Evaluation Online. *Journal of Communication*, *60*(3), 413–439. https://doi.org/10.1111/j.1460-2466.2010.01488.x

Metzger, M. J., Flanagin, A. J., Mena, P., Jiang, S., & Wilson, C. (2021). From Dark to Light: The Many Shades of Sharing Misinformation Online. *Media and Communication*, *Volume 9*(Issue 1), 134–143. https://doi.org/10.17645/mac.v9i1.3409

Milfont, T. L., & Duckitt, J. (2010). The environmental attitudes inventory: A valid and reliable measure to assess the structure of environmental attitudes. *Journal of Environmental Psychology*, *30*(1), 80–94. https://doi.org/10.1016/j.jenvp.2009.09.001

Milli, S., Belli, L., & Hardt, M. (2021). From Optimizing Engagement to Measuring Value. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 714–722. https://doi.org/10.1145/3442188.3445933

Molina, M. D., Wang, J., Sundar, S. S., Le, T., & DiRusso, C. (2023). Reading, Commenting and Sharing of Fake News: How Online Bandwagons and Bots Dictate User Engagement. *Communication Research*, *50*(6), 667–694. https://doi.org/10.1177/00936502211073398

Narayanan, A. (2023, March 9). *Understanding Social Media Recommendation Algorithms*. Knight First Amendment Institute. http://knightcolumbia.org/content/understanding-social-media-recommendation-algorithms

Nyhan, B., & Reifler, J. (2010). When Corrections Fail: The Persistence of Political Misperceptions. *Political Behavior*, *32*(2), 303–330. https://doi.org/10.1007/s11109-010-9112-2

Pang, N., Ho, S. S., Zhang, A. M. R., Ko, J. S. W., Low, W. X., & Tan, K. S. Y. (2016). Can spiral of silence and civility predict click speech on Facebook? *Computers in Human Behavior*, *64*, 898–905. https://doi.org/10.1016/j.chb.2016.07.066

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, *54*(4), 1643–1662. https://doi.org/10.3758/s13428-021-01694-3

Pennycook, G., Bear, A., Collins, E. T., & Rand, D. G. (2020). The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings. *Management Science*, *66*(11), 4944–4957. https://doi.org/10.1287/mnsc.2019.3478

Pennycook, G., Binnendyk, J., Newton, C., & Rand, D. G. (2021). A Practical Guide to Doing Behavioral Research on Fake News and Misinformation. *Collabra: Psychology*, *7*(1), 25293. https://doi.org/10.1525/collabra.25293

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, *592*(7855),

590–595. https://doi.org/10.1038/s41586-021-03344-2

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. *Psychological Science*, *31*(7), 770–780. https://doi.org/10.1177/0956797620939054

Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, *188*, 39–50. https://doi.org/10.1016/j.cognition.2018.06.011

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

Pew Research Center. (2024, September 17). Social Media and News Fact Sheet. *Pew Research Center*. https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/

Pierre, J., & Neuman, S. (2021, October 27). How decades of disinformation about fossil fuels halted U.S. climate policy. *NPR*. https://www.npr.org/2021/10/27/1047583610/once-again-the-u-s-has-failed-to-take-sweeping-climate-action-heres-why

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Poon, C. S. K., Koehler, D. J., & Buehler, R. (2014). On the psychology of self-prediction: Consideration of situational barriers to intended actions. *Judgment and Decision Making*, *9*(3), 207–225. https://doi.org/10.1017/S1930297500005763

Renault, T., Mosleh, M., & Rand, D. G. (2025). Republicans are flagged more often than Democrats for sharing misinformation on X's Community Notes. *Proceedings of the National Academy of Sciences*, *122*(25), e2502053122. https://doi.org/10.1073/pnas.2502053122

Sethi, P. (2024, April 22). *What are climate misinformation and disinformation and what is their impact?* Grantham Research Institute on Climate Change and the Environment. https://www.lse.ac.uk/granthaminstitute/explainers/what-are-climate-misinformation-and-disinformation/

Shin, S. Y., & Lee, J. (2022). The Effect of Deepfake Video on News Credibility and Corrective Influence of Cost-Based Knowledge about Deepfakes. *Digital Journalism*, *10*(3), 412–432. https://doi.org/10.1080/21670811.2022.2026797

Smelter, T. J., & Calvillo, D. P. (2020). Pictures and repeated exposure increase perceived accuracy of news headlines. *Applied Cognitive Psychology*, *34*(5), 1061–1071. https://doi.org/10.1002/acp.3684

Statistics Canada, G. of C. (2024, July 25). *The spread of misinformation: A multivariate analysis of the relationship between individual characteristics and fact-checking behaviours of Canadians*. https://www150.statcan.gc.ca/n1/pub/22-20-0001/222000012024003-eng.htm

Sundar, S. S. (2008). The MAIN Model: A Heuristic Approach to Understanding Technology

Effects on Credibility. *In: M. J. Metzger and A. J. Flanagin, Eds., Digital Media and Learning*, 73-100.

Sundar, S. S., Molina, M. D., & Cho, E. (2021). Seeing Is Believing: Is Video Modality More Powerful in Spreading Fake News via Online Messaging Apps? *Journal of Computer-Mediated Communication*, *26*(6), 301–319. https://doi.org/10.1093/jcmc/zmab010

Sundar, S. S., Snyder, E. C., Liao, M., Yin, J., Wang, J., & Chi, G. (2025). Sharing without clicking on news in social media. *Nature Human Behaviour*, *9*(1), 156–168. https://doi.org/10.1038/s41562-024-02067-4

Syrdal, H. A., & Briggs, E. (2018). Engagement with Social Media Content: A Qualitative Exploration. *Journal of Marketing Theory and Practice*, *26*(1–2), 4–22. https://doi.org/10.1080/10696679.2017.1389243

Tam, K.-P., & Chan, H.-W. (2023). Conspiracy theories and climate change: A systematic review. *Journal of Environmental Psychology*, *91*, 102129. https://doi.org/10.1016/j.jenvp.2023.102129

Todorov, A., Chaiken, S., & Henderson, M. (2002). The Heuristic-Systematic Model of Social Information Processing todorov—Recherche Google. In *The persuasion handbook: Developments in theory and practice* (p. pp.195-212). SAGE Publications, Inc. https://doi.org/10.4135/9781412976046

Trunfio, M., & Rossi, S. (2021). Conceptualising and measuring social media engagement: A systematic literature review. *Italian Journal of Marketing*, *2021*(3), 267–292. https://doi.org/10.1007/s43039-021-00035-8

Vaccari, C., & Chadwick, A. (2020). Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media + Society*, *6*(1). https://doi.org/10.1177/2056305120903408

Yadav, A., Phillips, M. M., Lundeberg, M. A., Koehler, M. J., Hilden, K., & Dirkin, K. H. (2011). If a picture is worth a thousand words is video worth a million? Differences in affective and cognitive processing of video and text cases. *Journal of Computing in Higher Education*, *23*(1), 15–37. https://doi.org/10.1007/s12528-011-9042-y

Zahn, M. (2025, July 1). Here's why Meta ended fact-checking, according to experts. *ABC News*. https://abcnews.go.com/US/why-did-meta-remove-fact-checkers-experts-explain/story?id=117417445

Appendices

A. Qualtrics Questionnaire (Pretest)

1. Introduction and Instructions

You will be presented with a video clip on a topic related to climate change. Please watch the entire video and answer the few questions that follow. We are interested in your opinion of the content.

After the study, we will ask a few questions about you. This is to help us understand the background of the participants in this study. Your responses will be kept confidential and will be used solely for research purposes.

⚠️ WARNING: The following video has not been fact-checked and may contain controversial or false information.

Please watch the video entirely and attentively with the sound on.

2. Stimuli

1. Glacier Retreat
2. CO_2 and Food Abundance
3. Debunking Hurricane Myths
4. The Challenges and Realities of Climate Modelling with Steven Koonin

3. Comprehension Questions

Video 1

According to the video, what was likely the coldest period of the last 10,000 years?

- Pleistocene glaciation (1)
- Cold War (2)
- Little Ice Age (3)
- Winters starting from 2001 (4)

Video 2

According to the video, what has the higher level of CO_2 done to plant life?

- It has made the Earth a lot greener (1)
- It has made the Earth less green (2)
- It hasn't changed how green the Earth is (3)

Video 3

What is this video about?

- Hurricanes are not as dangerous as tsunami (1)
- Climate change has NOT been making hurricanes stronger and more frequent (2)
- Climate change has been making hurricanes stronger and more frequent (3)
- Earthquakes are following the same trend than hurricanes (4)

Video 4

What point is the narrator of the video making?

- There is only one accurate climate model (1)
- Climate models and their predictions should be trusted (2)
- Climate models are not accurate enough to make trustworthy predictions (3)

4. Credibility Scale

On a scale of (1) very poorly to (7) very well, how well do the following adjectives describe the content you just saw?

- Accurate
- Believable
- Authentic

5. Behavioral Engagement Variable

Would you like more information on the topic you just learned about?

- No (1)
- Yes (2)

6. Engagement Scale

If you were to see this content on social media, how likely would you be to **share it on social media?**

- Extremely unlikely (1)
- Somewhat unlikely (2)
- Neither likely nor unlikely (3)
- Somewhat likely (4)
- Extremely likely (5)

If you were to see this content on social media, how likely would you be **to leave a reaction (e.g. a like or an emoji)?**

- Extremely unlikely (1)
- Somewhat unlikely (2)
- Neither likely nor unlikely (3)
- Somewhat likely (4)
- Extremely likely (5)

If you were to see this content on social media, how likely would you be to **leave a**

comment?

- Extremely unlikely (1)
- Somewhat unlikely (2)
- Neither likely nor unlikely (3)
- Somewhat likely (4)
- Extremely likely (5)

7. Controls

Have you encountered the information contained in this video clip before?

- No (1)
- Maybe (2)
- Yes (3)

8. Did you notice a warning regarding the content?

- No (1)
- Yes (2)

8. Content Quality

To which extent did you find the video interesting?

- Not interesting at all (1)
- Slightly interesting (2)
- Moderately interesting (3)
- Very interesting (4)
- Extremely interesting (5)

How would you rate the quality of the video you just watched?

- Terrible (1)
- Poor (2)

- Average (3)
- Good (4)
- Excellent (5)

How would you rate the quality of the audio from the video you just watched?

- Terrible (1)
- Poor (2)
- Average (3)
- Good (4)
- Excellent (5)

How would you rate the quality of the content organization from the video you just watched?

- Terrible (1)
- Poor (2)
- Average (3)
- Good (4)
- Excellent (5)

9. Demographics

What is your age range?

- Under 18 (1)
- 18 - 24 (2)
- 25 - 34 (3)
- 35 - 44 (4)
- 45 - 54 (5)
- 55 - 64 (6)
- 65 - 74 (7)
- 75 - 84 (8)

- 85 or older (9)
- Prefer not to say (10)

What is your gender?

- Male (1)
- Female (2)
- Non-binary / third gender (3)
- Prefer not to say (4)

How would you describe your political leaning?

- Strongly left leaning (1)
- Left leaning (2)
- Somewhat left leaning (3)
- Centre (4)
- Somewhat right leaning (5)
- Right leaning (6)
- Strongly right leaning (7)
- Prefer not to say (8)

What is the highest level of education you've completed?

- Less than high school (1)
- High school diploma or equivalent (GED) (2)
- Some college, no degree (3)
- Associate degree (4)
- Bachelor's degree (5)
- Master's degree (6)
- Doctoral degree (PhD) (7)
- Professional degree (e.g., MD, JD, MBA) (8)
- Prefer not to say (9)

10. Environmental Attitude Scale

On a scale of (1) Strongly disagree to (7) Strongly agree, how would you rate the following items?

a) I would like to join and actively participate in an environmentalist group.
b) I would NOT get involved in an environmentalist organization.
c) Modern science will NOT be able to solve our environmental problems.
d) Modern science will solve our environmental problems.
e) Humans are severely abusing the environment.
f) I do NOT believe that the environment has been severely abused by humans.

B. Questionnaire (Main Study)

1. Introduction

You will be presented with a [video clip/ audio clip/ text excerpt] on a topic related to climate change. Please watch/listen/read the entire thing and answer the few questions that follow. We are interested in your opinion of the content.

After the study, we will ask a few questions about you. This is to help us understand the background of the participants in this study. Your responses will be kept confidential and will be used solely for research purposes.

2. Instructions

You are about to watch/listen/read a [video clip/audio clip/text excerpt] that presents information on climate change. Once you're done, please continue to the next section to answer a few questions.

Warning: The [video/audio/text] has not been fact-checked and may contain controversial or false information.

Please watch/listen/read the [video clip/audio clip/text excerpt] entirely and attentively with the sound on.

https://www.youtube.com/watch?v=gckYBeX6_Vc&ab_channel=CO2Coalition

3. Comprehension Check

1. According to the video, what has the higher level of CO_2 done to plant life?

- It has made the Earth a lot greener (1)
- It has made the Earth less green (2)

- It hasn't changed how green the Earth is (3)

4. Credibility scale

How well do the following adjectives describe the content you just saw? (on a scale of 1-very poorly to 7- very well)

- Accurate
- Believable
- Authentic

5. Behavioral Engagement Variable

Would you like more information on the topic you just learned about?

- No (1)
- Yes (2)

6. Engagement Scale

If you were to see this content on social media, how likely would you be to **share it on social media?**

- Extremely unlikely (1)
- Moderately unlikely (2)
- Slightly unlikely (3)
- Neither likely nor unlikely (4)
- Slightly likely (5)
- Moderately likely (6)
- Extremely likely (7)

If you were to see this content on social media, how likely would you be **to leave a reaction (e.g. a like or an emoji)?**

- Extremely unlikely (1)

- Moderately unlikely (2)
- Slightly unlikely (3)
- Neither likely nor unlikely (4)
- Slightly likely (5)
- Moderately likely (6)
- Extremely likely (7)

Please select **Slightly likely** for this question to show you are paying attention.

- Extremely unlikely (1)
- Moderately unlikely (2)
- Slightly unlikely (3)
- Neither likely nor unlikely (4)
- Slightly likely (5)
- Moderately likely (6)
- Extremely likely (7)

If you were to see this content on social media, how likely would you be to **leave a comment ?**

- Extremely unlikely (1)
- Moderately unlikely (2)
- Slightly unlikely (3)
- Neither likely nor unlikely (4)
- Slightly likely (5)
- Moderately likely (6)
- Extremely likely (7)

7. Control

Have you encountered the information contained in this video clip before?

- No (1)

- Maybe (2)
- Yes (3)

8. Demographics

What is your age range?

- Under 18 (1)
- 18 - 24 (2)
- 25 - 34 (3)
- 35 - 44 (4)
- 45 - 54 (5)
- 55 - 64 (6)
- 65 - 74 (7)
- 75 - 84 (8)
- 85 or older (9)
- Prefer not to say (10)

What is your gender?

- Male (1)
- Female (2)
- Non-binary / third gender (3)
- Prefer not to say (4)

How would you describe your political leaning?

- Strongly left leaning (1)
- Left leaning (2)
- Somewhat left leaning (3)
- Centre (4)
- Somewhat right leaning (5)
- Right leaning (6)

- Strongly right leaning (7)
- Prefer not to say (8)

What is the highest level of education you've completed?

- Less than high school (1)
- High school diploma or equivalent (GED) (2)
- Some college, no degree (3)
- Associate degree (4)
- Bachelor's degree (5)
- Master's degree (6)
- Doctoral degree (PhD) (7)
- Professional degree (e.g., MD, JD, MBA) (8)
- Prefer not to say (9)

9. Environmental Attitude Scale

On a scale of (1) Strongly disagree to (7) Strongly agree, how would you rate the following items?

a) I would like to join and actively participate in an environmentalist group.
b) I would NOT get involved in an environmentalist organization.
c) Modern science will NOT be able to solve our environmental problems.
d) Modern science will solve our environmental problems.
e) Humans are severely abusing the environment.
f) I do NOT believe that the environment has been severely abused by humans.

10. Science Media Literacy Scale

On a scale of (1) Never to (6) Every time, how would you rate the following items?

a) I think about what point of view a science broadcaster or writer is trying to support.
b) I think about whether sources of science news have my best interest in mind.

c) I check to see if a science fact in a news story is backed up by credible sources.

d) I have changed my thinking about a science topic when I received new information.

C. Debrief (Pretest)

Thank you for taking the time to complete this study. We appreciate your participation.

The purpose of this study was to investigate which of the videos you saw is the most persuasive. All these videos have been identified as making dubious, misleading claims about climate change and came from organizations that have been labelled as promoting climate change disinformation. Disinformation is information that has been created with the intention to mislead people. This was not shared before you answered the questions as we didn't want to bias your answers.

Here are videos and articles on the topics you were exposed to, providing information supported by strong scientific consensus.

- Video 1: Glaciers Retreat
    - Climate 101: Glaciers | National Geographic
    - Ice sheets in Greenland, Antartica melting faster than previously thought, research shows
    - Antarctic glacier the size of Florida more vulnerable to warming than previously thought, experts warn
- Video 2: CO_2
    - Plants Are Struggling to Keep Up with Rising Carbon Dioxide Concentrations
    - CO_2: How an essential greenhouse gas is heating up the planet
    - Why CO_2 matters for climate change - BBC News
- Video 3: Hurricanes
    - How climate change makes hurricanes worse
    - How climate change is changing hurricanes
    - How climate change is making hurricanes more dangerous
- Video 4: Climate data models
    - How scientists calculate climate change

    - How to understand climate modelling – and why you should care | Shannon Algar | TEDxKingsPark
    - Study Confirms Climate Models are Getting Future Warming Projections Right

To learn more about how to identify and protect yourself from disinformation, you can visit the following websites:

- Climate Science Literacy Initiative by Meta, in partnership with Monash, Cambridge and Yale University
- Climate Change Fact Check by Factcheck.org
- Misinformation Resilience Toolkit by Poynter.org
- Climate Disinformation Database by DeSmog.org
- Climate Misinformation Myths by the Environmental Defense Fund

These resources provide tips on how to evaluate the environmental content that you see online.

Thank you again for your participation. If you have any questions about the study or wish to learn more, please feel free to contact us through Prolific or at the email addresses written below.

D. Debrief (Main Study)

Thank you for taking the time to complete this study. We appreciate your participation, which will allow us to better understand message credibility depending on media type.

More specifically, the purpose of this study is to investigate if a certain type of media (video, audio, or text) is more credible for content that promotes disinformation. Disinformation is information that has been created with the intention to mislead people. This was not shared before you answered the questions as we didn't want to bias your answers.

The content you were exposed to contained some dubious, misleading statements about climate change from an organization that have been identified as presenting such disinformation.

Here are videos and articles that share information that has shown strong consensus in the scientific community.

- Plants Are Struggling to Keep Up with Rising Carbon Dioxide Concentrations
- CO_2: How an essential greenhouse gas is heating up the planet
- Why CO_2 matters for climate change - BBC News

To learn more about how to identify and protect yourself from disinformation, you can visit the following websites. These resources provide tips on how to evaluate the environmental content that you see online.

- Climate Science Literacy Initiative by Meta, in partnership with Monash, Cambridge and Yale University
- Climate Change Fact Check by Factcheck.org
- Misinformation Resilience Toolkit by Poynter.org
- Climate Disinformation Database by DeSmog.org
- Climate Misinformation Myths by the Environmental Defense Fund

E. Credibility Scale (Appelman & Sundar, 2016)

How well do the following adjectives describe the content you just read (from 1 = describes very poorly to 7 = describes very well)?

- accurate
- authentic
- believable

F. Engagement Scale

For this scale, a structure similar to Pang et al.(2016) was adapted for the pretest and main study.

Pang et al., (2016)'s version

6-point Likert scale (from 1 = Highly unlikely to 6 = Highly likely)

- Intention to like main post or individual comments
    - How likely are you to click 'like' on the main Facebook post?
    - How likely are you to click 'like' on the individual comments?
- Intention to comment on story on Facebook
    - How likely are you to comment on the post?
- Intention to share story on Facebook
    - How likely are you to share the post on your own wall?

Pretest

On a scale of (1) Extremely unlikely to (5) Extremely likely, if you were to see this video on social media, how likely would you be to…

- leave a reaction (e.g., a like or an emoji)?
- share it on social media?
- leave a comment?

Main study

On a scale of (1) Extremely unlikely to (7) Extremely likely, if you were to see this video on social media, how likely would you be to…

- leave a reaction (e.g., a like or an emoji)?
- share it on social media?
- leave a comment?

G. Environmental Attitude Scale (Milfont & Duckitt, 2010)

For this research, we only selected the questions that are part of the brief version (†), for three of the dimensions that were most related to the topic of climate change.

- Dimension 1: Enjoyment of nature
    - 01. I am NOT the kind of person who loves spending time in wild, untamed wilderness areas. (R)
    - 02. I really like going on trips into the countryside, for example to forests or fields.*, †
    - 03. I find it very boring being out in wilderness areas. (R)*
    - 04. Sometimes when I am unhappy, I find comfort in nature. 05. Being out in nature is a great stress reducer for me.*
    - 06. I would rather spend my weekend in the city than in wilderness areas. (R)
    - 07. I enjoy spending time in natural settings just for the sake of being out in nature. 08. I have a sense of well-being in the silence of nature.*
    - 09. I find it more interesting in a shopping mall than out in the forest looking at trees and birds. (R)*
    - 10. I think spending time in nature is boring. (R)*, †

- Dimension 2: Support for interventionist conservation policies
    - 01. Industry should be required to use recycled materials even when this costs more than making the same products from new raw materials.
    - 02. Governments should control the rate at which raw materials are used to ensure that they last as long as possible.*, †
    - 03. Controls should be placed on industry to protect the environment from pollution, even if it means things will cost more.*
    - 04. People in developed societies are going to have to adopt a more conserving life-style in the future.*

- 05. The government should give generous financial support to research related to the development of alternative energy sources, such as solar energy.
- 06. I don't think people in developed societies are going to have to adopt a more conserving life-style in the future. (R)*
- 07. Industries should be able to use raw materials rather than recycled ones if this leads to lower prices and costs, even if it means the raw materials will eventually be used up. (R)*
- 08. It is wrong for governments to try and compel business and industry to put conservation before producing goods in the most efficient and cost effective manner. (R)
- 09. I am completely opposed to measures that would force industry to use recycled materials if this would make products more expensive. (R)
- 10. I am opposed to governments controlling and regulating the way raw materials are used in order to try and make them last longer. (R)*, †

- **Dimension 3: Environmental movement activism**
  - 01. If I ever get extra income I will donate some money to an environmental organization.
  - **02. I would like to join and actively participate in an environmentalist group.*, †**
  - 03. I don't think I would help to raise funds for environmental protection. (R)*
  - **04. I would NOT get involved in an environmentalist organization. (R)*, †**
  - 05. Environmental protection costs a lot of money. I am prepared to help out in a fund-raising effort.*
  - 06. I would not want to donate money to support an environmentalist cause. (R)*
  - 07. I would NOT go out of my way to help recycling campaigns. (R)
  - 08. I often try to persuade others that the environment is important.
  - 09. I would like to support an environmental organization.*
  - 10. I would never try to persuade others that environmental protection is important. (R)

- Dimension 4: Conservation motivated by anthropocentric concern
    - 01. One of the best things about recycling is that it saves money.
    - 02. The worst thing about the loss of the rain forest is that it will restrict the development of new medicines.
    - 03. One of the most important reasons to keep lakes and rivers clean is so that people have a place to enjoy water sports.*, †
    - 04. Nature is important because of what it can contribute to the pleasure and welfare of humans.*
    - 05. The thing that concerns me most about deforestation is that there will not be enough lumber for future generations.*
    - 06. We should protect the environment for the well being of plants and animals rather than for the welfare of humans. (R)
    - 07. Human happiness and human reproduction are less important than a healthy planet. (R)
    - 08. Conservation is important even if it lowers peoples' standard of living. (R)*
    - 09. We need to keep rivers and lakes clean in order to protect the environment, and NOT as places for people to enjoy water sports. (R)*, †
    - 10. We should protect the environment even if it means peoples' welfare will suffer.(R)*

- **Dimension 5: . Confidence in science and technology**
    - 01. Most environmental problems can be solved by applying more and better technology.
    - 02. Science and technology will eventually solve our problems with pollution, overpopulation, and diminishing resources.*
    - 03. Science and technology do as much environmental harm as good. (R)
    - **04. Modern science will NOT be able to solve our environmental problems. (R)*, †**

- 05. We cannot keep counting on science and technology to solve our environmental problems. (R)*
- 06. Humans will eventually learn how to solve all environmental problems.*
- 07. The belief that advances in science and technology can solve our environmental problems is completely wrong and misguided. (R)*
- 08. Humans will eventually learn enough about how nature works to be able to control it.
- 09. Science and technology cannot solve the grave threats to our environment. (R)
- **10. Modern science will solve our environmental problems.*, †**

- **Dimension 6: . Environmental threat**
  - 01. If things continue on their present course, we will soon experience a major ecological catastrophe.*
  - 02. The earth is like a spaceship with very limited room and resources.
  - 03. The balance of nature is very delicate and easily upset.
  - 04. When humans interfere with nature it often produces disastrous consequences.*
  - **05. Humans are severely abusing the environment.*, †**
  - 06. The idea that we will experience a major ecological catastrophe if things continue on their present course is misguided nonsense. (R)
  - 07. I cannot see any real environmental problems being created by rapid economic growth. It only creates benefits. (R)
  - 08. The idea that the balance of nature is terribly delicate and easily upset is much too pessimistic. (R)*
  - **09. I do not believe that the environment has been severely abused by humans. (R)*, †**
  - 10. People who say that the unrelenting exploitation of nature has driven us to the brink of ecological collapse are wrong. (R)*

- Dimension 7: Altering nature

- 01. Grass and weeds growing between paving stones may be untidy but are natural and should be left alone. (R)
- 02. The idea that natural areas should be maintained exactly as they are is silly, wasteful, and wrong.
- 03. I'd prefer a garden that is wild and natural to a well groomed and ordered one. (R)*, †
- 04. Human beings should not tamper with nature even when nature is uncomfortable and inconvenient for us. (R)*
- 05. Turning new unused land over to cultivation and agricultural development should be stopped. (R)*
- 06. I'd much prefer a garden that is well groomed and ordered to a wild and natural one.*, †
- 07. When nature is uncomfortable and inconvenient for humans we have every right to change and remake it to suit ourselves.*
- 08. Turning new unused land over to cultivation and agricultural development is positive and should be supported.
- 09. Grass and weeds growing between pavement stones really looks untidy.*
- 10. I oppose any removal of wilderness areas no matter how economically beneficial their development may be. (R)

- Dimension 8: Personal conservation behaviour
  - 01. I could not be bothered to save water or other natural resources.(R)*
  - 02. I make sure that during the winter the heating system in my room is not switched on too high.
  - 03. In my daily life I'm just not interested in trying to conserve water and/or power. (R)*
  - 04. Whenever possible, I take a short shower in order to conserve water.
  - 05. I always switch the light off when I don't need it on any more.*
  - 06. I drive whenever it suits me, even if it does pollute the atmosphere. (R)
  - 07. In my daily life I try to find ways to conserve water or power.*

    - 08. I am NOT the kind of person who makes efforts to conserve natural resources. (R)*, †
    - 09. Whenever possible, I try to save natural resources.*, †
    - 10. Even if public transportation was more efficient than it is, I would prefer to drive my car. (R)

- Dimension 9: Human dominance over nature
    - 01. Humans were meant to rule over the rest of nature.*
    - 02. Human beings were created or evolved to dominate the rest of nature.*, †
    - 03. Plants and animals have as much right as humans to exist. (R)*
    - 04. Plants and animals exist primarily to be used by humans.*
    - 05. Humans are as much a part of the ecosystem as other animals. (R)
    - 06. Humans are no more important in nature than other living things. (R)
    - 07. Nature exists primarily for human use.
    - 08. Nature in all its forms and manifestations should be controlled by humans.
    - 09. I DO NOT believe humans were created or evolved to dominate the rest of nature.(R)*, †
    - 10. Humans are no more important than any other species. (R)*

- Dimension 10: Human utilization of nature
    - 01. It is all right for humans to use nature as a resource for economic purposes.
    - 02. Protecting peoples' jobs is more important than protecting the environment.*, †
    - 03. Humans do NOT have the right to damage the environment just to get greater economic growth. (R)*
    - 04. People have been giving far too little attention to how human progress has been damaging the environment. (R)
    - 05. Protecting the environment is more important than protecting economic growth. (R)*
    - 06. We should no longer use nature as a resource for economic purposes. (R)

- 07. Protecting the environment is more important than protecting peoples' jobs. (R)*, †
- 08. In order to protect the environment, we need economic growth.
- 09. The question of the environment is secondary to economic growth.*
- 10. The benefits of modern consumer products are more important than the pollution that results from their production and use.*

- Dimension 11: Ecocentric concern
  - 01. The idea that nature is valuable for its own sake is naı̈ve and wrong. (R)*
  - 02. It makes me sad to see natural environments destroyed.
  - 03. Nature is valuable for its own sake.*
  - 04. One of the worst things about overpopulation is that many natural areas are getting destroyed.
  - 05. I do not believe protecting the environment is an important issue. (R)*
  - 06. Despite our special abilities humans are still subject to the laws of nature.*
  - 07. It makes me sad to see forests cleared for agriculture.*, †
  - 08. It does NOT make me sad to see natural environments destroyed. (R)*, †
  - 09. I do not believe nature is valuable for its own sake. (R)
  - 10. I don't get upset at the idea of forests being cleared for agriculture. (R)

- Dimension 12: Support for population growth policies
  - 01. We should strive for the goal of ''zero population growth''.
  - 02. The idea that we should control population growth is wrong. (R)
  - 03. Families should be encouraged to limit themselves to two children or less.*, †
  - 04. A married couple should have as many children as they wish, as long as they can adequately provide for them. (R)*, †
  - 05. Our government should educate people concerning the importance of having two children or less.*
  - 06. We should never put limits on the number of children a couple can have. (R)*
  - 07. People who say overpopulation is a problem are completely incorrect. (R)

- 08. The world would be better off if the population stopped growing.
- 09. We would be better off if we dramatically reduced the number of people on the Earth.*
- 10. The government has no right to require married couples to limit the number of children they can have. (R)*

Note. R = reversed coded items.

* The 72 balanced items selected for the short version of the EAI (i.e., EAI-S).

† The 24 balanced items selected for the brief version of the EAI (i.e., EAI-24).

H. Science Media Literacy (Austin et al., 2023)

For the main study, we selected the two most relevant items from each dimension of the scale, rated on a 6-point scale ranging from (1) Never to (6) Every time.

- Source
    - I check whether those who create science news know about the topic.
    - **I think about what point of view a science broadcaster or writer is trying to support.**
    - I look to see if those who share science news on social media have checked the accuracy of their facts.
    - **I think about whether sources of science news have my best interests in mind.**
    - I think about whether those who provide science information might be doing so to gain power or profit.
    - I get science news from multiple sources to make sure I get the full story.

- Content
    - I think about how scientists can draw different conclusions from the same science facts.
    - **I check to see if a science fact in a news story is backed up by a credible source.**
    - I check to see if a picture or graph accurately matches the scientific information it represents.
    - I check to see if the science news I read is up to date.
    - I think about whether a news story with real science facts could still lead to a false conclusion.
    - **I have changed my thinking about a science topic when I received new information.**