

NBER WORKING PAPER SERIES

TOWARD AN UNDERSTANDING OF THE ECONOMICS OF MISINFORMATION:
EVIDENCE FROM A DEMAND SIDE FIELD EXPERIMENT ON CRITICAL THINKING

John A. List
Lina M. Ramírez
Julia Seither
Jaime Unda
Beatriz Helena Vallejo

Working Paper 32367
<http://www.nber.org/papers/w32367>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2024

The opinions and conclusions expressed herein are solely those of the authors and should not be construed as representing the opinions or policies of the sponsoring agencies and research partners. We thank the financial services and insurance companies SURA and Proteccion for funding this project. We thank Gustavo Castillo and Maximilian Hippold for their excellent research assistance. This study was reviewed by the University of Chicago IRB and granted an exemption (IRB22-1217). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at <http://www.nber.org/papers/w32367>

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2024 by John A. List, Lina M. Ramírez, Julia Seither, Jaime Unda, and Beatriz Helena Vallejo. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Toward an Understanding of the Economics of Misinformation: Evidence from a Demand
Side Field Experiment on Critical Thinking

John A. List, Lina M. Ramírez, Julia Seither, Jaime Unda, and Beatriz Helena Vallejo

NBER Working Paper No. 32367

April 2024

JEL No. C93,D9,D91

ABSTRACT

Misinformation represents a vital threat to the societal fabric of modern economies. While the supply side of the misinformation market has begun to receive increased scrutiny, the demand side has received scant attention. We explore the demand for misinformation through the lens of augmenting critical thinking skills in a field experiment during the 2022 Presidential election in Colombia. Data from roughly 2,000 individuals suggest that our treatments enhance critical thinking, causing subjects to more carefully consider the truthfulness of potential misinformation. We furthermore provide evidence that reducing the demand of fake news can deliver on the dual goal of reducing the spread of fake news by encouraging reporting of misinformation.

John A. List
Department of Economics
University of Chicago
1126 East 59th
Chicago, IL 60637
and Australian National University
and also NBER
jlist@uchicago.edu

Jaime Unda
Ethos Behavioral Team
unda10a@gmail.com

Beatriz Helena Vallejo
Ethos Behavioral Team
beatriz.vallejo@ethosbt.com

Lina M. Ramírez
Department of Economics
University of Chicago
lmramirez@uchicago.edu

Julia Seither
Department of Economics
Universidad de Rosario
and JILAE, NOVAFRICA.
Bogota
Colombia
julia.seither@urosario.edu.co

1 Introduction

One would be hard-pressed to find an issue as divisive for modern nations as the spread of misinformation. Yet our understanding of the economic sources of misinformation remains nascent. Given that the spread of misinformation can negatively affect election outcomes (DellaVigna and Kaplan, 2007), health outcomes (Bursztyn et al., 2020), and increase polarization (Azzimonti and Fernandes, 2023), the foundation of modern societies is critically linked to people’s beliefs and their trust in institutions.

Through an economics lens, much of the work to date addressing misinformation relates to the supply side of the information market, by fact-checking news (Barrera et al., 2020) or using social media initiatives and advertising (Breza et al., 2021). This work has lent important insights into the causes and consequences of misinformation on the supply side. Yet, every market has two sides and to understand the market for misinformation and craft efficient policies to combat its spread, we also need to develop an understanding of the demand side.

Our study represents only a small step in this direction by exploring how nudges to critical thinking skills affect the demand and treatment of misinformation. We leverage the critical thinking framework of List (2022) to examine if treatments that teach individuals to "slow down" and reflect on their behavior makes them less vulnerable to misinformation. To operationalize our approach, we conduct an online field experiment to provide evidence relating the vulnerability to misinformation to critical thinking in a context with high probability of fake news prevalence - the Colombian Presidential elections in 2022. Colombia is severely politically polarized (Edelman, 2023), and although Colombia’s democracy is one of the longest-standing in Latin America, the country still experiences violence and human rights abuse. Reducing the demand for misinformation may have meaningful impacts on the functioning of democracy.

We conducted an online framed field experiment with 2,235 adults by randomizing individuals into one of four groups. The first group watches a thought-provoking video that shows how automatic thinking and misperceptions can affect our everyday lives. The second group completes a personality test that shows them their cognitive traits and how this

makes them prone to behavioral biases. A third group watches both the video and completes the personality test. We benchmark treatment effects against a control group without any intervention. Additionally, we explore four potential mediators in the relationship between our treatments and fake news vulnerability: trust, own discrimination perception, ambiguity aversion, and dehumanization (Rodríguez-Pérez et al., 2011; Webster and Kruglanski, 1994; Williams et al., 1997).

Our main outcomes of interest are misinformation vulnerability and stated willingness to report fake news on social media. Our experimental results show that the video interventions are successful in decreasing misinformation vulnerability. Individuals who watch the video are 30.3% (4 pp) less likely to consider fake news reliable, overall. Distinguishing between news types, we find that point estimates are larger for neutral than for political news, potentially speaking to a higher stickiness of political misinformation vulnerability. However, individuals in the control group are more likely to consider neutral news reliable than political news implying larger treatment impacts of videos on political misinformation. The effect is strongest for fake news targeting the political right. We do not find significant effects of the personality assessment alone and it provides no additional benefit to watching only the video.

Overall, these findings provide evidence that exposing individuals to their own behavioral biases is not as effective as videos that highlight relatable, sympathetic individuals with different backgrounds and beliefs. Interestingly, we find that the joint intervention (personality test and video) affects the perceptions about the reliability of true news headlines. We find that individuals who receive the joint intervention are 30.7% (4.4 pp) less likely to consider true headlines reliable. Neither the video intervention nor the personality test alone alters perceptions regarding the reliability of true news headlines overall or covering political news. However, the video intervention on its own does affect skepticism towards neutral news that is true.

This result suggests that while our video intervention fosters critical thinking by cultivating skepticism toward fake headlines and facilitating the differentiation between true and false political news, the tools provided by the video to enhance discernment are somewhat limited. When combined with the personality test, individuals generally become more skept-

tical but struggle to discern between what is genuine and what is fabricated. A promising avenue for future research lies in identifying tools that not only increase skepticism but also bolster discernment.

While the videos on their own decrease misinformation vulnerability, they don't seem to be sufficient to affect the spread of misinformation. Using our measure of reporting fake news tweets we do not find that videos increase tweet reporting. We find, however, that individuals who receive the joint intervention (video and behavioral assessment) are 15% (8.5 pp) more likely to state that they are willing to report the fake non-political tweet (Tweet A). This result suggests that decreasing misinformation vulnerability only leads to (stated) behavior changes if combined with personalized information about own behavioral biases.

Lastly, regarding our four potential mediators, we find sizable impacts of our treatments on dehumanization and discrimination perception but not on trust or ambiguity aversion. Watching the video increases the likelihood that a respondent states that they do not feel discriminated against by roughly 19.5% (4.5 pp), suggesting a reduction in perceived discrimination. We also find that watching the video decreases by 15% (10.1 pp) and 9% (6.1 pp) the use of negative characteristics to describe the political left and political right, respectively, suggesting a reduction in dehumanization. Again, we find that the personality tests on their own do not affect any of our four mediators.

Our results on the impact of video interventions are consistent with recent research on interventions targeting political polarization. For example, Voelkel et al. (2023) show that the most effective among 25 interventions to reduce political animosity and correct democratic misperception was a video depicting a series of interactions between pairs of people with different political orientations, but that are able to connect and find mutual respect.

We view our study as containing three key insights on the demand side. First, videos that provoke viewers to slow down and reconsider their thinking are successful in increasing skepticism and reducing misinformation vulnerability. Yet, our assessment that aims at a similar result by providing individuals with evidence about their cognitive traits has limited effects. Second, the joint intervention could potentially affect the spread of misinformation in that video-and-test-treated individuals are more likely to report a tweet as misinformation. This provides some hope that demand side factors can have spillover effects to the supply side

of this market. Third, we find suggestive evidence that the videos are effective because they encourage viewers to dehumanize less and to reflect more critically on their discrimination experiences.

The remainder of our study flows as follows. The next section details the context and experimental design. Section 3 describes our empirical strategy. Section 4 summarizes our main results and Section 5 concludes with a brief discussion of external validity.

2 Context and Experimental Design

2.1 Colombian Presidential Election in 2022

At the beginning of 2022 the retiring right-wing president of Colombia, Ivan Duque, had only 25% of the populations' approval. Results of the primaries in March showed that the center lacked support and that citizens were seeking radical change. After the primaries two opposite candidates emerged as favorites, polarizing the country: Rodolfo Hernandez, claiming to be independent but supported by the right-wing parties, and Gustavo Petro, a left-wing senator and former guerrilla member.

This polarized environment was perfect for misinformation to arise. Journalists in the region compared the 2022 elections to "a schoolyard food fight, noticing multiple trends in mis- and disinformation being slung from left, right and center". Fact checking outlets, like *Colombiacheck* saw an increase in new outlets spreading misinformation and a high volume of de-contextualized photos of candidates regardless of their political orientation

2.2 Experimental Design

In this context, Ethos Behavioral Team (Ethos BT) created the Detox Information Project¹ (DIP) a digital and information literacy project based in behavioral science. Its purpose is to provide tools to help people and organizations to be more critical of information. We partner with Ethos BT to investigate the efficacy of various tools aimed at bolstering critical thinking skills to mitigate vulnerability to misinformation. Following List (2022), our focus

¹See the project here: <https://www.somosdip.com/>.

lies predominantly on interventions designed to fortify one of critical thinking’s dual pillars: developing and assimilating empirical evidence and updating one’s beliefs. To do so, we run an online framed field experiment (Harrison and List, 2004) that explores three different tools: a thought-provoking video that shows how behavioral biases can affect our everyday lives, a personality test that shows how prone we are to automatic thinking and misperceptions, and a combination of the video and the personality test.

Furthermore, because we wish to test theory and speak to policymakers, we use "Option C thinking" (List, 2024). To do so, we examine whether four psychological traits identified by DIP as correlates of vulnerability to misinformation are mediators of the relationship between our interventions targeting critical thinking enhancement and our outcome of misinformation vulnerability. Van Bavel et al. (2021) argues that psychological risks, including political polarization, and the intertwining of emotionality and morality, are key factors affecting the path between the exposure to and sharing of misinformation. Based on this, DIP focused on identifying these psychological risks to design the interventions we evaluate in this paper. DIP ran a survey experiment with 1038 workers from five Colombian companies to identify psychological risks related to misinformation vulnerability. With these data, DIP identified four psychological traits associated with misinformation vulnerability: trust, dehumanization, discrimination perception, and ambiguity aversion.

Participants were randomized into one of three treatment groups or a control group without intervention. Those who were in the video only group were randomly exposed to one of four distinct videos that follow the same structure, but focus on a different mediator. The videos show interactions between real people² in a neutral setting. Each interaction brings together people from distinct social groups in Colombia. The first part of the video introduces each person to the viewer and shows how the participants get to know each other free of any identity attribution. In this interaction, they share personal stories that visibly sensitize the respective other(s). The videos then show sequences that were filmed before the interaction. In those sequences, each person makes statements about the other social group(s) – based on

²The protagonists of the videos were not actors. The videos can be found under the following links: Stereotyping/ambiguity video: <https://www.youtube.com/watch?v=eKbb5RE5cw0>; Discrimination video: <https://www.youtube.com/watch?v=wt8-XD5AAIs>; Trust video: <https://www.youtube.com/watch?v=oeK1lN4pkbI>; Dehumanization video: <https://www.youtube.com/watch?v=2TzbFrJFoxQ>.

societal stereotypes and stigmas. The video shows them watching these sequences together with their counterpart(s) and then reflecting on their former statements, revising them, and acknowledging that they made fast statements based on stereotypes. The videos encourage the viewer to question their own cognitive traits by "slowing down" their thinking.

Like the videos, there are four personality tests that follow the same structure, but each focuses on a different mediator. In each test, participants are asked to answer questions about themselves and their political preferences. Then, they are exposed to a set of statements that depend on the mediator we are testing. After completing the survey, respondents receive their test results. The results explain what the test is measuring and the respondent's biases and vulnerability to misinformation in terms of the four mediators.

With three treatments ($D = 3$) and four mediators ($M = 4$) we adopt a *partial* factorial design randomization. In total, we have $D \times M = 12$ treatment groups and one control group. Due to logistical constraints, we did not implement a full factorial design (16 treatments groups) including the interaction of all the mediators with the control group, meaning that the control group only responded to the survey but did not receive any placebo interventions.

2.3 Data

To elicit misinformation vulnerability, participants are presented with 19 distinct news headlines in random order and are prompted to determine their reliability as in Roozenbeek and Van der Linden (2019). These headlines are curated from articles reviewed by fact-checking platforms such as *ColombiaCheck* or *El Detector*, affiliated with the national news outlet *La Silla Vacía*, ensuring the reliability of their truthfulness. Additionally, we categorize the news articles based on their political ideology to explore potential correlations between susceptibility to misinformation and political inclinations. Among the 19 selected headlines, ten are fake and nine are authentic, or actually true, news. Of the fake headlines, four lean towards the left, four lean towards the right, and two are neutral. Conversely, among the true headlines, three lean towards the left, three lean towards the right, and three are neutral. News headlines were presented next to logos imitating news agencies (neither names nor logos actually exist in Colombia). This was to prompt participants to think about information coming from news outlets rather than social networks though to avoid deception we did not

explicitly link any headlines with logos.

Following the classification of the news headlines, participants are presented with two tweets and asked to assess their appropriateness and indicate if they would report them. Both tweets are fictitious, with one (Tweet A) devoid of political content, whereas the other (Tweet B) carries political connotations. Lastly, participants are prompted to respond to questions concerning our four potential mediators. Upon completing the survey, all participants receive a bonus to purchase an ice cream from a renowned ice cream shop with branches across the country.

The intervention was made using *Datasketch*, a Colombian firm similar to Qualtrics. Unfortunately, we were unable to track attrition rates as *Datasketch* only retained data from participants who completed the entire survey. Nonetheless, we can confidently affirm that those who completed the survey complied to their assigned treatment, as they were unable to fast-forward the video or skip questions in the personality test.

In total, 3297 individuals completed our field experiment. However, our final sample is composed of 2235 individuals after eliminating duplicates based on identifiers built upon national identification numbers and surnames. This exclusion criteria was applied to individuals who either duplicated their participation or shared their personalized links. Of those remaining, 645 individuals are in the control group, 750 are treated with the personality test only, 408 are treated with the video only, and 432 are treated with both the personality test and the video.

The mean age in our sample is 33.5 years and all subjects are eligible to vote in elections (over 18 years old). The sample is balanced in terms of gender, with 52.8% of the sample being women. In terms of education level, 37.8% have some level of school education, 30.1% have a technical degree, 16.1% have a professional degree, and 9.4% have a graduate degree. In terms of place of residence, 58.1% of the sample is living in Bogota, and the rest live in other regions of the country.

2.4 Sampling Strategy

We compose our sample by inviting individuals from the general Colombian population to participate in our experiment. The invitation mentioned a small prize to be won and learning

about fake news vulnerability. We attempted to be as generic in our call as possible to avoid experimenter demand effects. We pursued three primary dissemination channels: i) Leveraging social networks such as Facebook, LinkedIn, Twitter, and WhatsApp to reach a broad audience. ii) Collaborating with a prominent journalist and TV presenter renowned for their extensive and diverse following across Colombia. iii) Partnering with private enterprises to distribute the intervention among their employees, enhancing its reach and impact.

Our underlying population of interest is the Colombian population that has access to digital platforms and social media. Our sample predominantly consists of individuals residing in Bogotá, the capital city of Colombia. While we recognize the importance of obtaining a more diverse representation from across the country, insights gathered from Bogotá residents are still highly informative for addressing our research question. As of January 2022, Bogotá accounted for approximately 10-12% of the nation’s total population, with around 30% of its residents being from outside the city. Additionally, given Colombia’s centralized political system, many pivotal decisions are made in Bogotá, further underscoring the significance of understanding the behaviors and attitudes of its inhabitants.

3 Empirical Strategy

Our general empirical approach will be standard and follows the typical analysis of data when the assignment mechanism is controlled. Let Y_i for unit i be our outcome of interest, let $D_i = \{0, 1, 2, 3\}$ be our treatment variable, where $D_i = 0$ is the control group, $D_i = 1$ is the personality test only, $D_i = 2$ is the video only and $D_i = 3$ is the treatment with both the personality test and the video. We group the distinct videos and personality tests together into one category each for this analysis and do not distinguish between the different types of videos or personality tests.

Let $M_i = \{0, 1, 2, 3\}$ be our mediator variable, where $M_i = 1$ is ambiguity aversion, $M_i = 2$ is dehumanization, $M_i = 3$ is perceived discrimination and $M_i = 4$ is trust. Our treatments D_i can yield an effect directly on Y_i , but also they can yield an effect indirectly on Y_i through the direct effect of D_i on M_i , holding the direct contribution of D_i fixed.

Accordingly, we define the ATE as:

$$\tau = \mathbb{E}[Y_i(M_i(d), d), -Y_i(M_i(0), 0)]$$

Assuming a parametric functional form, we estimate the following regression:

$$Y_i = \beta_0 + \beta_1 D_{i1} + \beta_2 D_{i2} + \beta_3 D_{i3} + X_i' \gamma + \mu_i \quad (1)$$

where D_{id} is equivalent to $D_i = d$ and X_i is a vector of control variables including age, gender, education level and whether the subject is from Bogota or anywhere else in Colombia. We are interested in β_1 , β_2 and β_3 .

Since towards the end of our intervention we asked participants questions related to our four potential mediators, with our data we can also speak to the effect of D_i on M_i . In particular, because our treatment D_i is statistically independent of potential outcomes $Y_i(M_i(d), d)$ and potential mediators $M_i(d)$, we can estimate:

$$M_i = \alpha_0 + \alpha_1 D_{i1} + \alpha_2 D_{i2} + \alpha_3 D_{i3} + X_i' \delta + \nu_i \quad (2)$$

for each of our four mediators of interest. We are interested in the coefficients α_1 , α_2 and α_3 . X_i is the same vector of controls we use in Equation 1.

4 Results

We summarize our results in two sections. First, we present the main treatment effect estimates. This summary includes direct estimates on our outcome metrics. In addition, we explore how our treatments affect the supply side of misinformation via the report of fake news. Then, in section 4.2 we consider potential mechanisms more carefully.

4.1 Main Results

The video intervention effectively reduces the likelihood of participants deeming fake headlines as reliable. In Table 1, we show that results from estimating Equation 1 reveal that those exposed to the video were 30.3% (4 pp) less likely to view any fake news as trustworthy compared to the control group (column (1)). This effect is even more pronounced when considering neutral headlines, with individuals being 32.3% (7.4 pp) less likely to perceive neutral fake news as credible compared to a control mean of 23% (column (2)). Effects for political headlines are similarly large, decreasing the likelihood that news is considered reliable by 5.3pp compared to a control mean of 17.4% for right-leaning news and a control mean of 20.6% for left-leaning news.

The personality test alone neither yielded a significant impact on participants' perceptions of the truthfulness of fake headlines, nor did it provide a synergistic benefit when combined with the video intervention. Although the joint treatment did reduce vulnerability to misinformation, its effectiveness is slightly smaller compared to the video intervention for right-leaning and neutral news.

Table 1: News

	Fake Headlines				True Headlines			
	(1) All	(2) Neutral	(3) Right	(4) Left	(5) All	(6) Neutral	(7) Right	(8) Left
Test	0.014 (0.018)	-0.000 (0.022)	0.023 (0.020)	-0.019 (0.021)	0.014 (0.019)	0.011 (0.021)	-0.016 (0.022)	0.011 (0.023)
Video	-0.040** (0.019)	-0.074*** (0.024)	-0.053** (0.021)	-0.053** (0.023)	-0.020 (0.021)	-0.051** (0.023)	-0.042 (0.025)	0.005 (0.026)
Both	-0.039** (0.018)	-0.065*** (0.024)	-0.039* (0.021)	-0.062*** (0.023)	-0.044** (0.019)	-0.042* (0.022)	-0.073*** (0.024)	-0.028 (0.025)
Test = Video (p)	0.003	0.001	0.000	0.116	0.088	0.005	0.286	0.824
Control Mean	0.132	0.229	0.174	0.206	0.143	0.195	0.236	0.226
Observations	2235	2235	2235	2235	2235	2235	2235	2235
R-Squared	0.064	0.060	0.059	0.048	0.061	0.063	0.033	0.017

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notably, our video intervention is also effective in decreasing participants' likelihood to perceive neutral true headlines as reliable. It does not have statistically significant impacts on true political news reliability, however.

Further, the joint intervention (personality test and video) was effective in decreasing participants' likelihood to perceive overall true headlines as reliable, with the exception of left-leaning headlines. Those who received the joint intervention are 30.7% (4.4 percentage points) less likely to regard true headlines as trustworthy. This outcome underscores how our video intervention when combined with the personality assessment fosters critical thinking by instilling skepticism toward all headlines, yet it falls short in equipping individuals with the discernment to differentiate between authentic and fabricated information.

In Table 2, columns (1) and (3) assess participants' perception of the appropriateness of tweet A and tweet B, respectively. In columns (2) and (4), the focus shifts to participants' willingness to report tweet A and tweet B, respectively. Recall that both tweets are fake. Tweet A is non-political and states "Vaccine fraud is being uncovered. With more affected people doubly injected every day, when fall arrives panic will reign." Tweet B is political, noting "The President is writing what would be a fragment of the state of "internal commotion" throughout the national territory."

Table 2: Tweets

	Tweet A		Tweet B	
	(1) Inappropriate	(2) Report	(3) Inappropriate	(4) Report
Test	-0.021 (0.024)	0.022 (0.026)	-0.020 (0.024)	0.002 (0.027)
Video	0.025 (0.027)	0.044 (0.031)	-0.014 (0.028)	0.030 (0.031)
Both	-0.005 (0.027)	0.085*** (0.030)	0.020 (0.027)	0.037 (0.031)
p-value Test = Video	0.081	0.470	0.812	0.351
Control Mean	0.735	0.584	0.709	0.571
Observations	2235	2235	2235	2235
R-Squared	0.047	0.011	0.062	0.009

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We find no statistically significant effect of any of our interventions increasing the percentage of people who consider the tweets as inappropriate. We attribute this result to the

fact that the percentage of individuals in the control group who considered that the tweets were inappropriate was already very high (over 70% for both tweets), making the effect of the treatments small in magnitude and statistically insignificant. Yet, our findings do indicate that individuals exposed to the joint intervention (video and behavioral assessment) are 15% (8.5 percentage points) more inclined to express willingness to report Tweet A as fake. This result suggests that our intervention exerts some influence on reporting non-political fake news, while exhibiting no impact on reporting fake political news. This result should be more carefully explored in future work as it highlights that a demand side intervention can potentially have spillover effects to the supply side, or the spread of disinformation.

4.2 Potential Mechanisms

Table 3: Discrimination, Ambiguity and Trust

	Discrimination	Ambiguity Aversion	Trust	Dehumanization		
	(1)	(2)	(3)	Left (4)	Neutral (5)	Right (6)
Test	0.015 (0.023)	0.015 (0.018)	-0.00619 (0.0262)	-0.028 (0.025)	0.005 (0.025)	-0.004 (0.025)
Video	0.045* (0.027)	0.017 (0.022)	0.0241 (0.0306)	-0.101*** (0.031)	-0.020 (0.029)	-0.061** (0.030)
Both	0.035 (0.027)	0.029 (0.022)	0.0197 (0.0301)	-0.081*** (0.030)	-0.038 (0.029)	-0.058** (0.029)
Test = Video (p)	0.254	0.917	0.309	0.015	0.387	0.047
Control Mean	0.231	0.122	0.597	0.674	0.676	0.693
Observations	2235	2235	2235	2235	2235	2235
R-Squared	0.025	0.006	0.021	0.026	0.048	0.039

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Results from estimating Equation 2 are reported in Table 3. Column (1) examines whether participants perceive discrimination based on income, race, gender, or social strata in their daily lives. Column (2), examines whether participants feel discomfort with uncertain situations (ambiguity aversion). Column (3) examines whether participants trust others in general. Lastly, Columns (4) to (6) gauge whether participants report negative emotions, such as recklessness, shame, or suspicion, towards voters of specific political views.

Our analysis reveals no significant impact on trust towards others or ambiguity aversion from any of our interventions. However, the video leads to a 19.5% (4.5 percentage points) increase in respondents stating they do not feel discriminated against, indicating a reduction in perceived discrimination. Additionally, watching the video leads to a decrease in the use of negative emotions to characterize both the political left and right by 15% (10.1 percentage points) and 9% (6.1 percentage points), respectively, suggesting a reduction in dehumanization.

The personality tests alone show no significant effect on any of the four mediators studied, and when combined with the video intervention, they provide no additional benefit. Overall, therefore, we find some evidence that the proposed mediators in the literature have some power as a causal pathway. We view this result as merely a first step in understanding the complete relationship between potential interventions and key outcome metrics around misinformation, but a promising one.

5 Conclusion

The causes and consequences of misinformation has become a key concern across governments, organizations, and scientists. For example, whether misinformation spreads like wild-fire or dies quickly on the vine merits serious consideration. And, understanding the factors that moderate that spread is invaluable. In this study, we depart from a traditional supply side investigation by conducting a field experiment that focuses on the demand side. In particular, we focus on enhancing critical thinking skills in an effort to alter the demand side in this market.

Our work yields several insights. First, our intervention is able to affect the level of individual discernment. This is good news. Yet, that discernment spills over to other types of news: whether fake or real, our subjects are more skeptical of any news. Second, we find that our intervention, in turn, potentially affects the spread of misinformation in that our treated individuals are more likely to report a tweet as misinformation. This provides a silver lining in that reducing the demand of fake news can deliver on the dual goal of reducing the spread of fake news by encouraging reporting of misinformation. Third, we find interesting

mediation paths, as we report suggestive evidence that the videos are effective because they encourage viewers to dehumanize less and to reflect more critically on their discrimination experiences.

In closing, we would be remiss not to mention the generalizability of our results and future research directions. In terms of the former, we follow List (2020) SANS conditions to understand the generalizability of these results. Our sample is a selected subset of Colombian individuals who agreed to participate in an online experiment. In this sense, whether the results generalize to other Colombian's, much less non-Colombian citizens is an open question. In terms of attrition, once a subject started very few opted out. Considering the naturalness of the choice task, setting, and time frame, we use a framed field experiment; thus, our setting is one in which individuals are engaged in a natural task and margin. Finally, in terms of scaling our insights, the interventions are simple and easy to administer so we argue that the benefit-cost profile should be enhanced at scale as the program expands. Finally, since we view our demand side results as WAVE1 insights, in the nomenclature of List (2020), replications need to be completed to understand if our results can be applied to other populations.

This point naturally leads to future research directions. Beyond replications, we urge researchers to more fully explore the suite demand side interventions in this market. This naturally begins with experimenting with several types of approaches, from pecuniary to non-pecuniary interventions, as well as the moderators and mediators of those effects. In this spirit, we trust that scholars will engage in creating causal moderators rather than stop at merely descriptive variants. Such explorations will help to detail what works and why. In addition, we trust that future work will tackle the interplay of the demand and supply sides of the misinformation market. Until key features are understood about the production and consumption sides, and relevant interactions, efficient policies to combat the creation and spread of misinformation will be difficult, if not impossible.

References

- Azzimonti, M. and Fernandes, M. (2023). Social media networks, fake news, and polarization. *European journal of political economy*, 76:102256.
- Barrera, O., Guriev, S., Henry, E., and Zhuravskaya, E. (2020). Facts, alternative facts, and fact checking in times of post-truth politics. *Journal of Public Economics*, 182:104123.
- Breza, E., Stanford, F. C., Alsan, M., Alsan, B., Banerjee, A., Chandrasekhar, A. G., Eichmeyer, S., Glushko, T., Goldsmith-Pinkham, P., Holland, K., et al. (2021). Effects of a large-scale social media advertising campaign on holiday travel and covid-19 infections: a cluster randomized controlled trial. *Nature medicine*, 27(9):1622–1628.
- Bursztyn, L., Rao, A., Roth, C. P., and Yanagizawa-Drott, D. H. (2020). Misinformation during a pandemic. Technical report, National Bureau of Economic Research.
- DellaVigna, S. and Kaplan, E. (2007). The fox news effect: Media bias and voting. *The Quarterly Journal of Economics*, 122(3):1187–1234.
- Edelman (2023). Edelman Trust Barometer Global Report 2023.
- Harrison, G. W. and List, J. A. (2004). Field experiments. *Journal of Economic literature*, 42(4):1009–1055.
- List, J. A. (2020). Non est disputandum de generalizability? a glimpse into the external validity trial. Technical report, National Bureau of Economic Research.
- List, J. A. (2022). Enhancing critical thinking skill formation: Getting fast thinkers to slow down. *The Journal of Economic Education*, 53(1):100–108.
- List, J. A. (2024). Optimally generate policy-based evidence before scaling. *Nature*, 626(10102):491–499.
- Rodríguez-Pérez, A., Delgado-Rodríguez, N., Betancor-Rodríguez, V., Leyens, J.-P., and Vaes, J. (2011). Infra-humanization of outgroups throughout the world. the role of similarity, intergroup friendship, knowledge of the outgroup, and status. *anales de psicología*, 27(3):679–687.

- Roozenbeek, J. and Van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, 5(1):1–10.
- Van Bavel, J. J., Harris, E. A., Pärnamets, P., Rathje, S., Doell, K. C., and Tucker, J. A. (2021). Political psychology in the digital (mis) information age: A model of news belief and sharing. *Social Issues and Policy Review*, 15(1):84–113.
- Voelkel, J. G., Stagnaro, M., Chu, J., Pink, S., Mernyk, J., Redekopp, C., Ghezze, I., Cashman, M., Adjodah, D., Allen, L., et al. (2023). Megastudy identifying effective interventions to strengthen americans’ democratic attitudes.
- Webster, D. M. and Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of personality and social psychology*, 67(6):1049.
- Williams, D. R., Yu, Y., Jackson, J. S., and Anderson, N. B. (1997). Racial differences in physical and mental health: Socio-economic status, stress and discrimination. *Journal of health psychology*, 2(3):335–351.