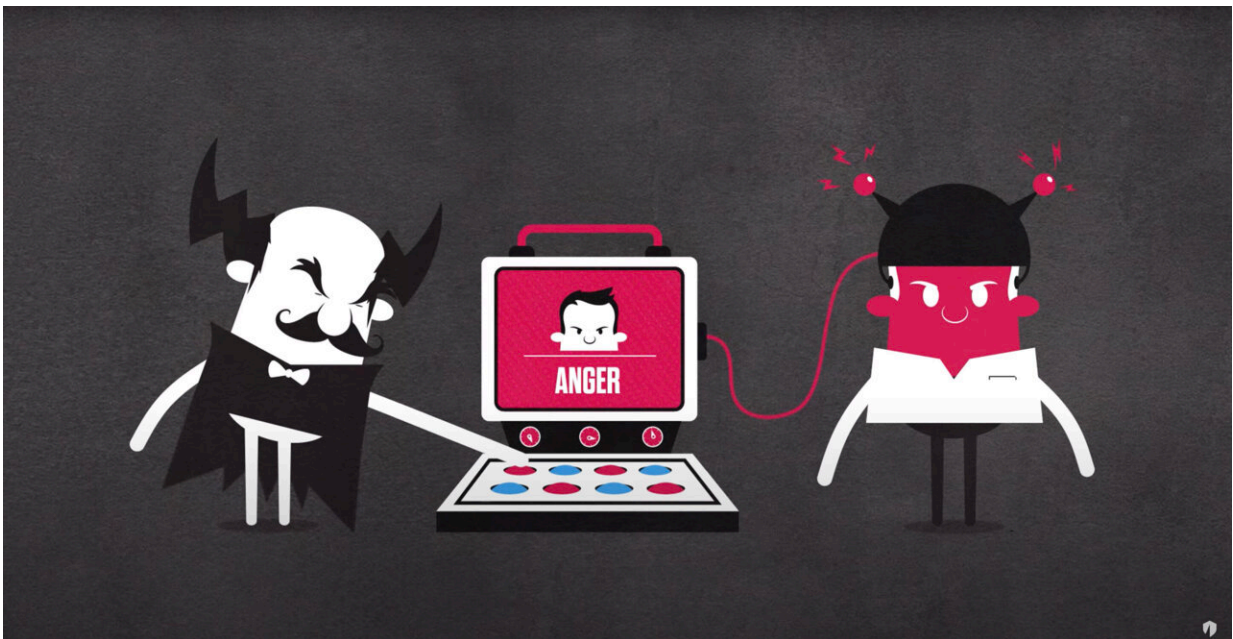


Experiment on YouTube reveals potential to 'inoculate' millions of users against misinformation

August 24 2022



A screenshot from the Inoculation Science video explaining Emotional Language use in misinformation. Credit: Inoculation Science

Short animations giving viewers a taste of the tactics behind misinformation can help to inoculate people against harmful content on social media when deployed in YouTube's advert slot, according to a major online experiment led by the University of Cambridge.

Working with Jigsaw, a unit within Google dedicated to tackling threats to open societies, a team of psychologists from the universities of Cambridge and Bristol created 90-second clips designed to familiarize users with manipulation techniques such as scapegoating and deliberate incoherence.

This "pre-bunking" strategy pre-emptively exposes people to tropes at the root of malicious propaganda, so they can better identify online falsehoods regardless of subject matter.

Researchers behind the [Inoculation Science](#) project compare it to a vaccine: by giving people a "micro-dose" of misinformation in advance, it helps prevent them falling for it in future—an idea based on what social psychologist's call "inoculation theory."

The findings, published in *Science Advances*, come from seven experiments involving a total of almost 30,000 participants—including the first "real world field study" of inoculation theory on a [social media](#) platform—and show a single viewing of a film clip increases awareness of misinformation.

The videos introduce concepts from the "misinformation playbook," illustrated with relatable examples from film and TV such as Family Guy or, in the case of false dichotomies, Star Wars ("Only a Sith deals in absolutes").

"YouTube has well over 2 billion active users worldwide. Our videos could easily be embedded within the [ad space](#) on YouTube to prebunk misinformation," said study co-author Prof Sander van der Linden, Head of the Social Decision-Making Lab (SDML) at Cambridge, which led the work.

"Our research provides the necessary proof of concept that the principle

of psychological inoculation can readily be scaled across hundreds of millions of users worldwide."

Lead author Dr. Jon Roozenbeek from Cambridge's SDML describes the team's videos as "source agnostic," avoiding biases people have about where information is from, and how it chimes—or not—with what they already believe.

"Our interventions make no claims about what is true or a fact, which is often disputed. They are effective for anyone who does not appreciate being manipulated," he said.

"The inoculation effect was consistent across liberals and conservatives. It worked for people with different levels of education, and different personality types. This is the basis of a general inoculation against misinformation."

Google—YouTube's parent company—is already harnessing the findings. At the end of August, Jigsaw will roll out a prebunking campaign across several platforms in Poland, Slovakia, and the Czech Republic to get ahead of emerging disinformation relating to Ukrainian refugees. The campaign is designed to build resilience to harmful anti-refugee narratives, in partnership with local NGOs, fact checkers, academics, and disinformation experts.

"Harmful misinformation takes many forms, but the manipulative tactics and narratives are often repeated and can therefore be predicted," said Beth Goldberg, co-author and Head of Research and Development for Google's Jigsaw unit.

"Teaching people about techniques like ad-hominem attacks that set out to manipulate them can help build resilience to believing and spreading misinformation in the future.

"We've shown that [video](#) ads as a delivery method of prebunking messages can be used to reach millions of people, potentially before harmful narratives take hold," Goldberg said.

The team argue that pre-bunking may be more effective at fighting the misinformation deluge than fact-checking each untruth after it spreads—the classic "debunk"—which is impossible to do at scale, and can entrench conspiracy theories by feeling like personal attacks to those who believe them.

"Propaganda, lies and misdirections are nearly always created from the same playbook," said co-author Prof Stephan Lewandowsky from the University of Bristol. "We developed the videos by analyzing the rhetoric of demagogues, who deal in scapegoating and false dichotomies."



A screenshot from the Inoculation Science video explaining use of scapegoating in misinformation. Credit: Inoculation Science

"Fact-checkers can only rebut a fraction of the falsehoods circulating online. We need to teach people to recognize the misinformation playbook, so they understand when they are being misled."

Six initial controlled experiments featured 6,464 participants, with the sixth experiment conducted a year after the first five to ensure earlier findings could be replicated.

Data collection for each participant was comprehensive, from basic information—gender, age, education, political leanings—to levels of numeracy, conspiratorial thinking, news and social media checking, "bullshit receptivity," and a personality inventory, among other "variables."

Factoring all this in, the team found that inoculation videos improved people's ability to spot misinformation, and boosted their confidence in being able to do so again. The clips also improve the quality of "sharing decisions": whether or not to spread damaging content.



A screenshot from the Inoculation Science video explaining use of emotional language in misinformation. Credit: Inoculation Science

Two of the animations were then tested "in the wild" as part of a vast experiment on YouTube, with clips positioned in the pre-video advert slot that provides an option to skip after five seconds.

Google Jigsaw exposed around 5.4 million US YouTubers to an inoculation video, with almost a million watching for at least 30 seconds. The platform then gave a random 30% of users that watched a voluntary test question within 24 hours of their initial viewing.

The clips aimed to inoculate against misinformation tactics of hyper-emotive language and use of false dichotomies, and the questions—based on fictional posts—tested for detection of these tropes. YouTube also gave a "control" group of users who had not viewed a video the same test question. In total, 22,632 users answered a question.

Despite the intense "noise" and distractions on YouTube, ability to recognize manipulation techniques at the heart of misinformation increased by 5% on average.

Google say the unprecedented nature of the experiment means there is no direct data comparison available. However, increases in brand awareness from advertising on YouTube—known as "brand lift"—are [typically limited to 1%](#) in surveys of under 45,000 users.

"Users participated in tests around 18 hours on average after watching the videos, so the inoculation appears to have stuck," said van der Linden.

Researchers say that such a recognition increase could be game changing if dramatically scaled up across social platforms—something that would be cheap to do. The average cost for each view of significant length was the tiny sum of US\$0.05.

Added Roozenbeek: "If anyone wants to pay for a YouTube campaign that measurably reduces susceptibility to misinformation across millions of users, they can do so, and at a miniscule cost per view."

First six experiments

The first six controlled experiments involved randomly assigning each participant with a 90-second "inoculation" video or a neutral control video. Participants were then randomly shown ten social media posts: five using deliberately manipulative techniques (although not all featured proven falsehoods), and five neutral posts. Participants were asked to rank levels of trust in the information, the degree to which they feel it is manipulative, and how likely they would be to share it.

Findings include:

- Emotional language video: "inoculated" participants were between 1.5 and 1.67 times better than the control group at identifying this manipulation technique.
- False dichotomies video: "inoculated" participants were 1.95—so almost twice—as good as the control group at identifying this manipulation technique.
- Incoherence video: "inoculated" participants were over twice as good (2.14) as the control group at identifying this manipulation technique.

YouTube experiment

The YouTube inoculation ad campaign ran over fifteen days in [YEAR] and targeted English-speaking users in the U.S. who were aged 18 years or over and had watched at least one political or news video on the platform.

In total, 22,632 participants answered a test question on YouTube: 11,432 who had seen an inoculation video, and 11,200 who had not.

Example questions as presented by YouTube:

- False dichotomy: Evaluate this sentence: "We either need to improve our education system or deal with crime on the streets." Users were asked to choose whether the sentence contained: a command; fearmongering; false dichotomy; none of these.
- Emotional language: Evaluate this sentence: "Baby formula linked to outbreak of new terrifying disease among helpless infants—parents despair." Users were asked to choose whether the sentence contained: a command; emotional language; false dichotomy; none of these.

More information: Jon Roozenbeek, Psychological inoculation

improves resilience against misinformation on social media, *Science Advances* (2022). DOI: [10.1126/sciadv.abo6254](https://doi.org/10.1126/sciadv.abo6254).
www.science.org/doi/10.1126/sciadv.abo6254

All the inoculation videos, along with background information on the approach, can be found at: <https://inoculation.science/>

Provided by University of Cambridge

Citation: Experiment on YouTube reveals potential to 'inoculate' millions of users against misinformation (2022, August 24) retrieved 24 April 2024 from <https://phys.org/news/2022-08-youtube-reveals-potential-inoculate-millions.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.