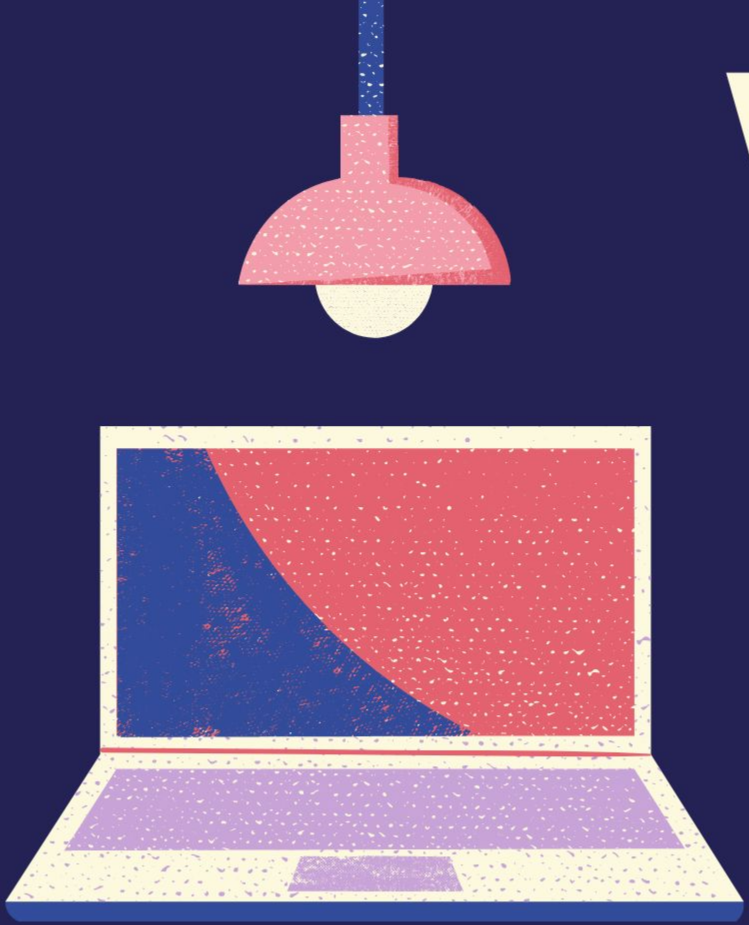
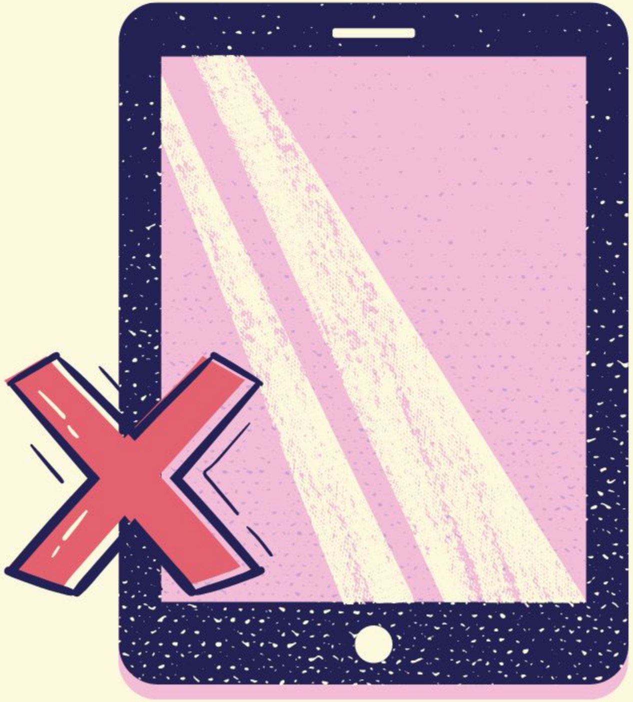


What do you Remember & Believe?

Exploring whether warning labels are effective in preventing the spread of misinformation when applied to posts shared by those we agree with



This post contains unverified information.

In an attempt to curb the spread of misinformation, social media platforms have begun implementing warning labels that indicate potentially unverified or disputed information within a post.

The study investigated whether memory and plausibility perceptions of misinformation online may be effectively suppressed by content warning labels when viewers' political agreement with the presenter of information is taken into account.

Which way do you lean?

19 Participants provided their self-reported political alignment and were presented with two fictional politicians, one of whom they opposed, and one who aligned with their views

2 Politicians, 3 Phases, Too Many Tweets.

In Phase 1, **the Tweets were presented randomly by either of these politicians,** and either with or without a warning label that disclaimed the content's possible misinformation.

Phase 2 presented participants with the same stimuli as well as random new statements, out of Tweet format.

Phase 3 presented rephrased versions of the stimuli, out of Tweet format.

Between stages, Participants were assessed on recall + asked to rank plausibility.

The Results.

Participants' **memory was worse for Tweets with warning labels**, and when they disagreed with the politician posting the tweet. The interaction between these made memory best in no-warning label conditions where participants agreed with the politician.

Participants rated plausibility lowest when there was a warning label AND when they disagreed with the politician posting the Tweet.

The Verdict?

Warning labels seem effective in deterring memory and plausibility perceptions for misinformation online, and this remains even when one agrees with the presenter of information.