



Master's thesis

*Mood, Threat, and Gamified Psychological
 Inoculation Against Misinformation*

Name and initials: Loughnan, D. P.

Student number: S4945778

E-mail address: d.p.loughnan@student.rug.nl

First assessor: Kai Epstude

Second assessor: Sabine Otten

Programme: Research Master Behavioural and Social Sciences

Theme: Understanding Societal Change

ECs: 30

Date: 10-7-2023

Word count:

9	9	7	3
---	---	---	---

Are there deviations of the Master's thesis from the proposed plan?

No Yes, please explain below the deviations

- In Study 1 the scale operationalising the DV was unreliable. As such, analyses were conducted for individual items and considered exploratory, for the purpose of informing the design of Study 2. Study 1 was thus a pilot study.
- Informed by the pilot study, the game Go Viral! was preferred because it is a shorter game than Bad News, and its effect on information discernment was unclear. In addition, in the main study threat was measured instead of manipulated, a control condition added, the sample size increased to allow power for a conditional process analysis, and the neutral mood group removed.
- The submission date was extended to 10th July, in consultation with assessors, due to a longer than anticipated EC committee request process for Study 2.

LLM Declaration: ChatGPT was occasionally used to clarify error messages in R, and less occasionally to suggest replacement syntax. No LLM was employed toward any other purpose, and thus did not contribute in any way to the writing of this thesis.

A thesis is an aptitude test for students. The approval of the thesis is proof that the student has sufficient research and reporting skills to graduate but does not guarantee the quality of the research and the results of the research as such, and the thesis is therefore not necessarily suitable to be used as an academic source to refer to. If you would like to know more about the research discussed in this thesis and any publications based on it, to which you could refer, please contact the supervisor mentioned.

Abstract

‘Inoculation games’ such as *Go Viral!* aim to ‘prebunk’ manipulative techniques of persuasion via psychological inoculation that: (i) instills a motivating perception of threat; and (ii) preemptively refutes an attack message. Stylistically similar ‘casual video games’ have been shown to promote happiness. But the Affect Infusion Model predicts information processing effects of a happy mood that could disrupt inoculation. Still, disruptive effects may be extinguished by sufficient motivation, such as that evoked by inoculation’s first step. We ran a preregistered, between-subjects, randomised controlled experiment on *Prolific* over two phases (UK residents, $N = 368$, 204 females, mean age 41.71 years). Participants underwent happy or sad mood inductions then either played *Go Viral!* or watched a control video. Happy mood decreased, and sad mood increased, discernment between reliable and unreliable news headlines. *Go Viral!* reduced ratings of unreliable news headlines, but results for reliability discernment favoured the control condition. Motivational threat was not higher in the treatment group. Results suggest that *Go Viral!* may not inoculate. This is the first study to jointly consider mood and threat in inoculation games, and the first to experimentally assess effects of mood states on susceptibility to online misinformation.

Keywords: psychological inoculation, prebunking, misinformation, mood, affect, threat

Mood, Threat, and Gamified Psychological Inoculation Against Misinformation

The broad acceptance of misinformation, defined as inaccurate information that is deliberately produced and intentionally or unintentionally propagated (Wu et al., 2019), has given rise to the notion of a “post-truth” world in which personal beliefs and appeals to emotion have more influence in shaping public opinion than objective facts (McIntyre, 2018). Relatedly, there has been in a recent rise in science denial (Iyengar & Massey, 2019), which is associated with outcomes such reduced vaccine uptake, hampering efforts to contain disease at the population level (Godlee et al., 2011; Kata, 2010; Loomba et al., 2021). In the context of the recent COVID-19 pandemic, the World Health Organisation (WHO) invoked the term “Infodemic” to characterise the effects of misinformation on public debate and understanding (WHO, 2020). Online misinformation is also linked to climate change denial (Cook et al., 2016; McCright et al., 2016), civic violence (Jolley & Paterson, 2020), and the disruption of processes vital to well-functioning democracies (Lewandowsky, Cook, et al., 2017; Lewandowsky, Lloyd, et al., 2017).

In today’s online environment there are considerable challenges to debunking, a common initial approach to combating misinformation. Firstly, the efficacy of a debunk may hinge on there being a clear and verifiable alternative explanation, which often there is not (Lewandowsky et al., 2012). Secondly, to debunk a misleading claim it must be invoked, which risks inciting the ‘illusory truth effect’ by which repeated claims are more likely to be judged true, regardless of veracity (Dechêne et al., 2010). A third challenge is the continued influence of misinformation, by which people rely on information initially taken as true even after it comes to be accepted as false (Chan et al., 2017; Walter & Tukachinsky, 2020). Finally, misinformation has been shown to travel further, faster, wider, and deeper in online social media than information verified as

true, and as such, efforts to retrospectively address it are routinely outpaced and overwhelmed (Vosoughi et al., 2018).

One solution to the issues associated with debunking is to pre-emptively refute, or ‘prebunk’ misinformation. Promising misinformation interventions that incorporate prebunking are ‘inoculation games’: choice-based video games set in simulated social media environments which aim to confer resistance to persuasion from manipulative techniques of persuasion common to misinformation, via a process of ‘psychological inoculation’ (Compton, 2013; Lewandowsky & van der Linden, 2021; McGuire, 1964). While early studies suggest the games are effective in reducing susceptibility to misinformation, several questions remain outstanding. These include what role affect may play in susceptibility to online misinformation, and in psychological inoculation (Compton et al., 2022; Pennycook et al., 2019). The question is pertinent in the context of games that inoculate given that stylistically similar ‘casual video games’ have been shown to substantially influence mood (Pine et al., 2020). Also, the role of the mood-related constructs of perceived threat in the outcomes of such games are largely unexplored despite the central role sense of threat holds in inoculation theorising. Finally, a recent reanalysis of data pertaining to inoculation games suggests that increased general scepticism, rather than better discernment for reliable and unreliable information, is the typical outcome from playing them (Modirrousta-Galian & Higham, 2022). This finding begs questions about what inoculation to misinformation should properly look like (Guay et al., 2022).

This preregistered experimental study will thus enquire into the role of mood in susceptibility to online misinformation, and into the role of both mood and perceived threat on the effects of inoculation game Go Viral!. In addition, it will compare post-intervention mean group ratings of unreliable information, the common metric employed in studies of inoculation games, with SDT measures of scepticism and discernment for both reliable and unreliable

information. The preregistration, R syntax, and clean and raw datasets are accessible at <https://osf.io/2t6fr/>.

Psychological Inoculation

Psychological inoculation theory (Compton, 2013; McGuire, 1964, 1970) makes an analogy to biological inoculation. Just as medical vaccines confer resistance to infection by providing exposure to attenuated strains of pathogens which prime an immune response to produce disease-fighting antibodies, psychological inoculation confers resistance to persuasion by providing exposure to a weakened persuasive message which stimulates a cognitive response to produce message-attacking counterarguments (Compton, 2013). Inoculation comprises two steps: (i) a warning of an impending threat to a preferred belief to motivate a subject to protect the belief; and (ii) a pre-emptive refutation of a weakened version of an anticipated persuasive attack to stimulate the production of counterarguments in associative memory (Compton, 2013; McGuire, 1964). By definition, psychological inoculation only occurs where this two-step process is described (Compton & Pfau, 2005; McGuire, 1964; Traberg et al., 2022).

Affect and Threat in Inoculation

Essential to the process of inoculation is the perception that there is an impending threat from a persuasive communication to some preferred belief that would thus be rendered vulnerable (Compton & Pfau, 2005). The sense of threat is what motivates the protection of an attitudinal position, and thus drives the production of counterarguments that is the ostensible outcome of inoculation (Compton, 2013). Threat has been identified as ‘the most distinguished feature of inoculation’ (Pfau, 1997, p. 137), and it is considered impossible to inoculate without it (Compton & Pfau, 2005). Indeed, a test for perceived threat has served as a manipulation check for inoculation over several decades (Banas & Richards, 2017). The measure traditionally used enquires into affective experiences of threat such as fear, anxiety, and perceived danger

elicited from a forewarning of a persuasive attack. Banas and Richards (2017) thus coined it ‘apprehensive threat’, and developed an alternate scale of ‘motivational threat’ to measure the extent to which a person is motivated to resist attitudinal change and counterargue persuasive assertions (see also Richards & Banas, 2018). Motivational threat, they argue, is more closely aligned to perceived threat featured in the first step of inoculation described by the theory, and is thus more appropriate for use in inoculation research. In testing the role of both types of threat in mediating relationships between an inoculation intervention and messages expounding 9/11 conspiracy theories, they found inoculation occurred only indirectly, and only via motivational threat. The relationships between both types of perceived threat and affect has in part led to suggestions that affect may play an important but as yet unelaborated role in inoculation (Compton et al., 2022; Fanselow, 2018).

That there may be factors beyond motivation and cognition at work in psychological inoculation, including a role for affect, has been mooted since the early days of the theory (Compton et al., 2022; Insko, 1968; Pfau et al., 2001). Recent evidence that apprehensive and motivational threat may be entangled with both mood and inoculation comes from Ivanov et al. (2020) who found higher levels of both types of threat in inoculated individuals, in contrast to the findings of Banas and Richards (2017), and also lower levels of happiness and higher levels of sadness. While this was the first study to consider sadness in inoculation, previous research found that negative-affect-eliciting forewarnings and inoculation messages led to both an increased sense of threat, and greater attitudinal resistance (i.e., inoculation; Miller et al., 2013; Pfau et al., 2009).

Mood and Judgement

Theories of mood and judgement also inform an argument that affect may play a meaningful role in psychological inoculation. Mood in this context is a mild, global, relatively

enduring affective state that is largely devoid of cognition or any salient cause, as distinct from emotion which is more intense, less enduring, has a cognitive component, and tends to arise from known antecedents (Forgas, 1995). Past research has established that happy and sad mood influence truth bias such that those in a happy mood are more likely to form favourable, positive judgements from ambiguous information (Forgas & East, 2008; McCornack & Parks, 1986). The influence of mood on judgement is theorised to come from two types of effect: informational and processing (Forgas, 2013, 2019).

Informational Effects

Informational effects include affect priming (Bower, 1981) by which mood states may selectively prime constructs in associative memory that are congruent with that state, and an affect-as-information effect (Schwarz, 1990) by which individuals may misattribute their current mood state as the informative outcome of their interaction with a piece of information. Both of these effects tend to lead people in a happy mood to form positive associations with information, and for those in a sad mood, negative associations. The Affect Infusion Model (AIM; Forgas, 1995, 2002), predicts that processing style will dictate which informational effect predominates. In terms of Bless and Fielder's (2006) Assimilative/Accommodative Processing Model, an accommodative, bottom-up processing style will be associated with affect priming, and an assimilative, top-down style associated with the affect-as-information model (Forgas, 1995, 2002).

Processing Effects

Regarding processing effects, the accommodative style is associated with more externally oriented and elaborative processing, better discernment, and a sad mood. Conversely, an assimilative style is associated with more internally oriented heuristic processing, poorer discernment between reliable and unreliable information, and happy mood. Using a signal

detection theory (SDT) approach, Forgas and East (2008) showed the outcomes of veracity judgements in happy and sad participants were consistent with the expected outcomes of informational and processing effects in that happy subjects were less sceptical and less accurate than sad. However, a key prediction of the AIM is that mood is not the sole determinant of processing style, and therefore does not drive processing and informational effects independent of other factors (Forgas, 2002). Specifically, the presence of motivating pressures may be sufficient to extinguish informational and processing effects of mood.

Mood and Inoculation

Research on mood, persuasion, and motivation suggests there may be a direct link between mood and motivation in the creation of counterarguments, the key mechanism of inoculation. Across four experiments, Forgas (2007) analysed the persuasive qualities of arguments produced by happy and sad participants, then tested their persuasive efficacy. Results showed that sad participants produced higher quality arguments that featured more concrete and tangible information, and which subsequently resulted in more real-life attitude change in others. Moreover, Experiment 4 included a motivating reward for highly persuasive messages, which saw the effect of mood substantially reduced. If mood is implicated in inoculation, the relevance of this research is apparent given the goal of inoculation is the production of high-quality counterarguments, and that the provision of adequate motivation is a necessary first step. Specifically, the presence of a processing effect of mood during inoculation would render happier people less likely to elaborate on attack messages and thus less likely to produce effective counterarguments to them, and an informational effect would render happier people less likely to perceive motivational warnings as truly threatening, thus thwarting the inoculation process at the outset.

Mood states are relevant also to inoculation games in particular. Such games emulate casual video games, which recent clinical research has shown are effective mood repair interventions (for a meta analysis see Pine et al., 2020; Rupp et al., 2017). That is, current inoculation games, or future games based on their scientific principles, may actively promote a happy mood. Thus, the influences of mood and motivation in this domain must be well understood so that they may be contained or optimised, as appropriate, lest the active ingredient of the remedy be neutralised by the method of delivery.

Inoculation Games

The need for new, scalable approaches to effectively confer resistance to online misinformation inspired a suite of inoculation games which differ in several ways to traditional inoculation interventions. For example, the games seek to protect not against specific arguments, but against techniques of manipulative persuasion that misinformative arguments commonly employ (e.g., the use of emotional language; Cook et al., 2017). This was a necessary innovation for prebunking misinformation because the content of misinforming arguments are often impossible to know in advance, whereas the presence of such techniques is predictable (Lewandowsky & van der Linden, 2021). Another departure from the majority of inoculation interventions is that these games broach highly contested issues, as they must in application to topics typically plagued by misinformation. Classic inoculation studies focused on cultural truisms, such as the benefits of regularly brushing one's teeth – the antithesis to contested issues. A third departure from many inoculation interventions is that the warning step is left implicit. The rationale for this is that perceived threat may be instilled not by directly referencing risks associated with an impending misinforming attack, but by the game's provision of mounting examples of ways one may be misinformed.

The first inoculation game was *Bad News* (Roozenbeek & van der Linden, 2019a, 2019b; van der Linden & Roozenbeek, 2020). In a Twittersphere-type environment, players take on the role of a news media tycoon seeking to maximise engagement, and thus their influence, by means of sensationalised misinformation. Gameplay is choice-based, and advances through six levels, each pertaining to a manipulative technique of persuasion. The general aim is to attract followers while maintaining sufficient credibility. *Bad News* has provided a template for several other games, including *Go Viral!* (Basol et al., 2021), which addresses misinformation specific to COVID-19. *Go Viral!* closely follows the structure and style of *Bad News* but incorporates only three techniques and is thus shorter (5 minutes compared to 15). Figure 1 compares elements of both games.

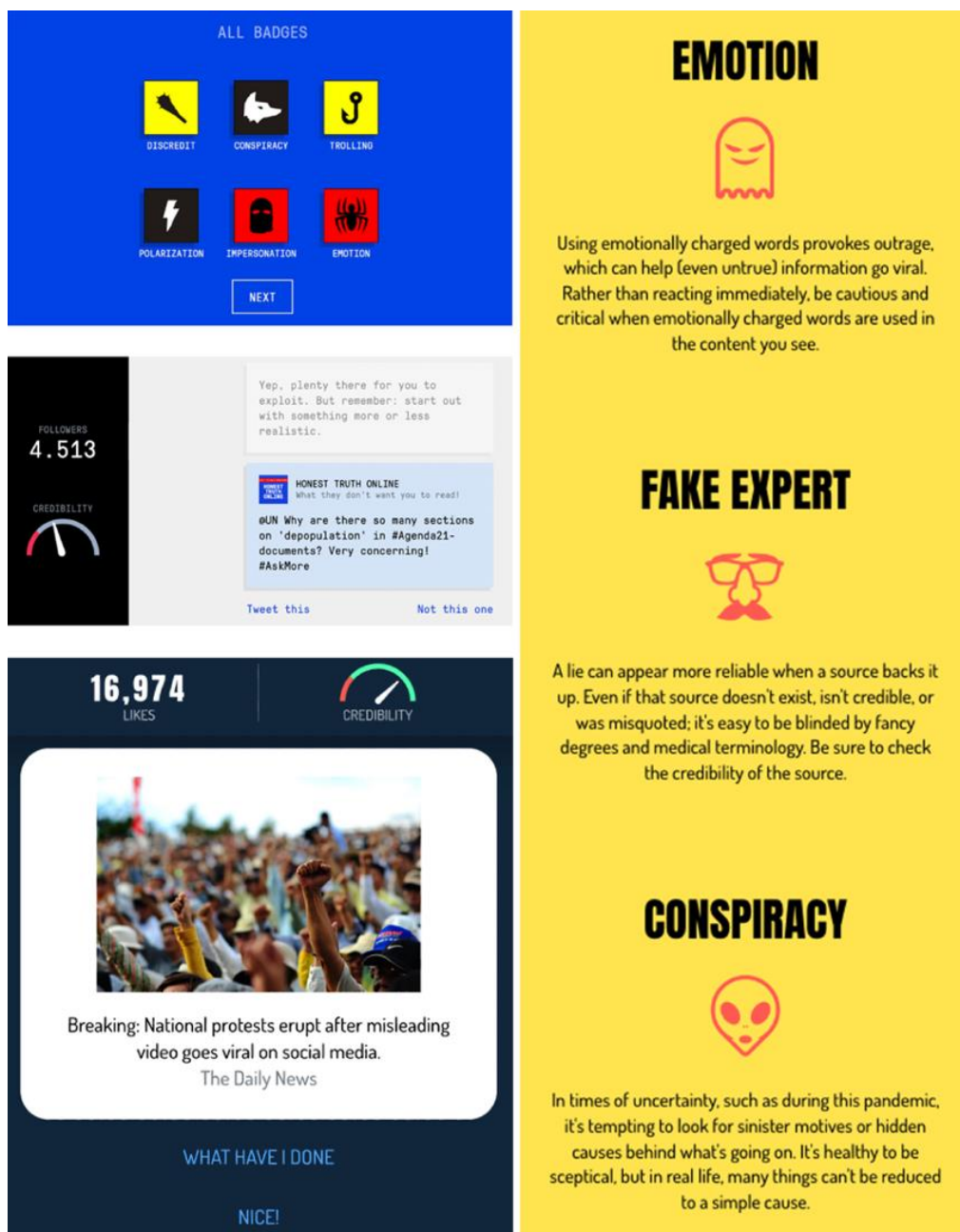
Paradigms for Assessing Effects on Susceptibility to Misinformation

Scales and Experimental Approach

Due to recent developments and discussions on the topic of inoculation games and how to best test for their effects on susceptibility to misinformation, what approach to take is a pertinent question for any study assessing such a game. The typical testing paradigm involves within-subjects pre- and post-tests on identical short texts that resemble social media posts or news headlines, and that either do (unreliable, ‘fake’) or do not (reliable, ‘real’) include manipulative techniques of persuasion (Roozenbeek & van der Linden, 2019a). This method has been shown to be relatively free of testing and item effects (Roozenbeek et al., 2021). However, until the development of the Misinformation Susceptibility Test (MIST; Maertens, Götz, et al., 2021), designed to reliably assess susceptibility to online misinformation, the internal consistency of the scales used were not reported, introducing uncertainty for the internal validity of aggregate measures.

Figure 1

Screenshots demonstrating elements of Bad News and Go Viral! gameplay



Note. From *Bad News*, the top left panel shows the ‘badges’ (levels) players collect, and the panel below shows a choice presented in the ‘conspiracy’ section. From *Go Viral!*, the bottom left frame demonstrates the close similarities in scoring, and the frame on the right the three levels of the shorter game, ‘fake expert’ (*Go Viral!*) being a type of ‘impersonation’ (*Bad News*).

Discernment, Scepticism, and the Statistical Approach

Most inoculation game studies to date have focused on mean group ratings of unreliable items only, and included reliable items solely to achieve some amount of balance between the two during testing. A live discussion in the field revolves around whether such ratings, or ratings of discernment and scepticism for both reliable and unreliable items, provides a better measure of ability to recognise manipulative techniques of persuasion and thus reduce susceptibility to misinformation by the theorised effect (Guay et al., 2022; Maertens, Götz, et al., 2021). Study 1 of Basol et al. (2021) did consider discernment and scepticism, reporting better discernment between reliable and unreliable information from playing Go Viral!. Discernment calculations in this case considered mean group difference-in-differences between pre-intervention scores on test items, and post-intervention scores on the same. Regarding scepticism, Basol et al., (Study 1; 2021) looked at responses to reliable news items only and reported no change.

Modirrousta-Galian & Higham (2022) reanalysed data from Basol et al. (2021), as well as four other studies that each assessed Bad News. They took a receiver operating characteristic approach from SDT to determine sensitivity (d' ; discernment) and response bias (c ; scepticism) for reliable and unreliable test items combined. A SDT approach does not assume linearity in the relationship between correct answers for reliable news items (the hit rate) and incorrect answers for unreliable (the false alarm rate), so is robust in this case where responses on reliable and unreliable items describe different distributions. In contrast, an analysis of mean ratings could find discernment where none exists (Higham et al., 2016). By the SDT method, Modirrousta-Galian & Higham (2022) calculated the increase in discernment from Study 1 of Basol et al. (2021) to be much smaller, signalling only anecdotal evidence for it. They also reported a general increase in scepticism for those who played Go Viral!. For Study 2 they found no better discernment, and again, increased scepticism. A general conclusion from their multiple

reanalyses was that the typical outcome of playing an inoculation game is increased scepticism for both reliable and unreliable test items, but no better discernment between the two. For Go Viral! specifically, however, they determined the impact on reliability discernment was unclear, and urged further research to clarify the effect.

Research Questions and Aims

The current study is a partial, conceptual replication of Basol et al. (Study 2; 2021), with the addition of induced mood states. It will focus on Go Viral! in addressing primary research questions relating to mood and threat in the context of inoculation games and susceptibility to online misinformation. This will allow also for the incorporation of secondary research questions related to relative effects of discernment and scepticism inferred from a SDT analytical approach, compared to those inferred by the more common metric of susceptibility to misinformation derived from mean group ratings of unreliable information.

The primary questions the current study will address are: does mood influence susceptibility to online misinformation; does mood bolster or thwart inoculation to misinformation conferred by an inoculation game, and; to what extent is perceived threat conferred by mood, and by playing an inoculation game? Secondary questions are: what are the effects of playing Go Viral! on discernment and scepticism for reliable and unreliable information, and; do apprehensive or motivational threat mediate the relationship between inoculation and susceptibility to misinformation, and does mood play a moderating role?

In addressing the research questions, this study will seek to use an internally consistent measure of susceptibility to online misinformation. It will examine what effect an induced mood state at the time of playing Go Viral! and during post-intervention testing may have, and at a later timepoint once the induced mood has dissipated. Also of central interest will be the effects

of mood on discernment and scepticism for reliable and unreliable news headlines, and on the post-intervention presence of apprehensive and motivational perceived threat.

Hypotheses

Immediately post-intervention (T2):

1. The inoculation condition will judge fake news headlines to be less reliable than will the control condition.
2. Compared to the sad mood group, the happy mood group will be:
 - a) less sceptical of reliable and unreliable news headlines combined.
 - b) less able to discern between reliable and unreliable news news headlines.
3. Compared to the control condition, the inoculation condition will perceive more:
 - a) apprehensive threat.
 - b) motivational threat.
4. Compared to the sad group that played Go Viral!, the happy group that played will perceive less:
 - a) apprehensive threat.
 - b) motivational threat.

At follow-up, one day later (T3):

5. The inoculation condition will judge unreliable news headlines to be less reliable than will the control condition.
6. Compared to the sad group that played Go Viral!, the happy group that played will judge unreliable news headlines to be more reliable.

Follow-up and Planned Exploratory Analyses

Discernment and scepticism for reliable and unreliable news headlines between the inoculation and control conditions will be assessed at T2 and T3. It is expected that there will be no difference in sensitivity, but the inoculation condition will be more sceptical.

Discernment and scepticism for reliable and unreliable news headlines between the happy and sad conditions will be assessed at T3. It is expected that there will be no difference in sensitivity, but the sad group will be more sceptical.

A mediation analyses will assess any mediating role of T2 apprehensive or motivational perceived threat on reliability ratings of news headlines at T3, in those who played Go Viral!. A mediating role is expected for both. Where there is mediation, a moderating role for mood will be explored via a conditional process analysis.

Method

Pilot Study

Preliminary research questions concerning main effects of an inoculation game and of mood on susceptibility to misinformation were broached within a preregistered pilot study. Details and analyses appear in Appendix A. Results provided preliminary support for a positive effect on accuracy of ratings of unreliable news headlines for the inoculation game Bad News, while there was partial support for an effect of sadness in reducing susceptibility to misinformation. The pilot study informed design decisions for the main study.

Operationalising Susceptibility to Misinformation

For the main study it was decided to use of the full 20-item Misinformation Susceptibility Test (MIST-20) instead of the 8-item scale (MIST-8), which was found to have unacceptable reliability with a considerable ceiling effect for item 4. It was in fact the low internal consistency of the MIST-8 that resulted in this first study serving as a pilot only, with analyses considered

more exploratory and thus often contrasted to those pre-registered. Further, due to confusion in one participant and a suspicion there were misunderstandings for others also, it was decided that contributors to the main study would be asked to classify news headlines as ‘reliable’ or ‘unreliable’ instead of ‘real’ or ‘fake’, classifications adopted here also in reporting. The reason for this is that test instructions do not make it sufficiently clear that ‘real’ and ‘fake’ refer to whether the statements are likely true, ‘yes’ or ‘no’, and not rather whether one might expect to encounter the headline the real world, or if one finds it in agreement with an already held belief of one’s own, or of a public figure that commonly makes ‘fake news’ determinations. Such framing effects have indeed been shown to substantially affect the psychometric properties of the MIST-20 (Roozenbeek et al., 2022), and analysis of the term ‘fake news’ has found it ambiguous at best (Habgood-Coote, 2019).

Control Condition

A non-gameplay control condition was chosen due to descriptive differences in post-intervention mood in the gameplaying control condition, consistent with a mood repair effect of casual video games (Pine et al., 2020). The subsequent control was to attend to a short video presenting text from a UN Innovation Network brief describing methods that organisations might employ in addressing misinformation. A control of this kind has the advantage of matching the intervention regarding the topic evoked, but without inoculating or providing any practical insights as to how an individual might protect themselves. It is also related to a comparison condition featured in Study 2 of Basol et al. (2021) that involved reading misinformation infographics from UNESCO, except those materials did indeed seek to confer protection to individuals.

Other Design Decisions

In favour of a shorter main study survey, decisions were also taken to forgo the inclusion of a neutral mood group, a test of social desirability, and a depressive symptoms screening, all of which lent no utility to the pilot study. Finally, a dimensional measure of mood validated for use over two timepoints, without the need to repeat items, was preferred.

Participants and Design

The present (main) study was a pre-registered, between-subjects, 2 (treatment, control) x 2 (happy, sad) double-blinded, randomised controlled factorial design experiment. It featured mood inductions and was carried out over two phases, one day apart. The between-subjects independent variables were inoculation conditions (the treatment Go Viral! and the control a non-inoculating, non-teaching video on related subject matter), and mood (induced happiness or sadness). Dependent variables were susceptibility to misinformation pre/post-intervention (T1/T2) and one day later (T3), and post-intervention perceived threat (apprehensive and motivational; T2). Age was included as a potential covariate. The Ethics Committee of the Faculty of Behavioural and Social Sciences at the University of Groningen granted ethics approval: PSY-2223-S-0395.

UK-residents between 18 and 68 years of age on the crowdsourcing platform *Prolific* made up the sample. A Monte Carlo simulation using estimated correlations between factors within a conditional process model completed in R (Donnelly et al., 2022) estimated the sample size required for an exploratory analysis of mediating and moderating effects for perceived threat and mood, respectively. It returned a sample size of 160 to provide 80% power in detecting a mediating effect of threat on the relationship between playing the game and headline ratings, beyond an estimated moderating influence of mood ($\alpha = .05$). As that size sample would be required in the treatment condition only, it was determined that $N = 320$ would be required to complete both phases of the experiment.

The study recruited and surveyed participants over three blocks¹. In total, 381 participants completed phase one. Data pertaining to 13 were removed: two did not provide consent, one did not generate item-level data, and 10 failed attention checks. This left $N = 368$. Group sizes at phase one by condition were $n_{\text{HAPPY-TREAT.}} = 89$, $n_{\text{HAPPY-CONT.}} = 93$, $n_{\text{SAD-TREAT.}} = 93$, and $n_{\text{SAD-CONT.}} = 93$. Of those invited to phase two, 350 returned and successfully completed the survey (5% attrition). Group sizes at phase two by condition were $n_{\text{HAPPY-TREAT.}} = 86$, $n_{\text{HAPPY-CONT.}} = 90$, $n_{\text{SAD-TREAT.}} = 85$, and $n_{\text{SAD-CONT.}} = 86$).

Materials

All questionnaires employed and described below appear in Appendix B, along with the original scales where there were adaptations. As pre-registered, preference was given to McDonald's Omega over Cronbach's Alpha as an assessment of internal consistency as the latter tends to underestimate reliability (McDonald, 1999; Revelle & Condon, 2019).

Susceptibility to Misinformation

The full 20-item version of the Misinformation Susceptibility Test (2023 adaptation; MIST-20; Maertens et al., 2021) operationalised the dependent variable, with 10 items being 'fake news', or unreliable headlines (i.e., that include a manipulative technique of persuasion; $\text{MIST}_{\text{FAKE}}$), and 10 being 'real news', or reliable news headlines (i.e., that do not include a manipulative technique of persuasion; $\text{MIST}_{\text{REAL}}$). Items were presented in random order. Ratings were dichotomous. Reliability of the full MIST-20 was acceptable ($\omega = .71$; $\omega_{\text{REAL}} = .70$ [acceptable]; $\omega_{\text{FAKE}} = .69$ [questionable]).

¹ An initial sample of 50 provided a pilot run to ensure the smooth technical flow of the survey and data collection procedures. The identification of two issues, one that allowed two participants on mobile devices to skip the consent page and another that inadvertently delayed another two participants on a survey page, led to changes and the recruitment of another block of 50 to assess fixes. No issues were apparent for block two, and the third block included the remainder of the sample. An estimation of phase two attrition and rates of rejected data in the first two blocks informed the number recruited for the final block.

Mood

The Multidimensional Mood State Questionnaire (MDMQ; Steyer et al., 1997) is an English translation of the original German scale. Response options were adapted from the suggested translations to more natural English language expressions. Mood was measured at two timepoints, employing items from left (T1) and right (T2) sides of the MDMQ. Items pertaining to the good-bad mood dimension only were selected, resulting in four items from each side, balanced for valence of mood. Responses were given on a 6-point Likert-type scale from ‘Not at all’ to ‘Very’, with those for negative emotions reverse coded so higher scores indicated happier mood. Reliability for the MDMQ was excellent ($\omega = .91$; $\omega_{\text{LEFT}} = .86$ [good]; $\omega_{\text{RIGHT}} = .93$ [excellent]).

Apprehensive Threat

The classic scale of perceived threat historically used in psychological inoculation research (Burgoon et al., 1978) constitutes a measure of apprehensive threat in the current study. The original scale presents five polar adjectives, for example, ‘non-threatening’ and ‘threatening’, which anchor each end of a 7-point response scale, the numbers 2-6 denoting matters of degree. These are proposed reactions to encountering a highly persuasive message that refutes a belief held by the respondent. In the case of Go Viral!, the focus is on manipulative techniques of persuasion rather than specific ideas or beliefs. As such, the scale used in the current study features adaptations to the statement to make it applicable to such techniques. Reliability for the scale was excellent ($\omega = .95$).

Motivational Threat

On the motivational threat scale (Banas & Richards, 2017), respondents are asked to register on a 7-point Likert-type scale their level of agreement, from ‘strongly disagree’ to ‘strongly agree’, with four statements pertaining to motivations to protect certain attitudes and beliefs related to the 9/11 attack. For the current study, statements were adapted to make them applicable to techniques of manipulative persuasion. Reliability of the motivational threat scale was acceptable ($\omega = .74$).

Moon Induction Procedures (MIPs)

To induce mood, the present study employed MIPs validated for online use by Marcusson-Clavertz et al. (2019). Each MIP begins with an instruction to adopt the target mood, then a 4-minute video followed by 4 minutes of music with an instruction to close one’s eyes and listen.

The sad mood induction included a clip from the animated film “The Lion King”. It begins with a wildebeest stampede and Simba, a lion cub, in danger. Simba’s father, Mufasa, saves him, but is drawn into the stampede. He crawls up to his brother, Scar, for help. But Scar allows Mufasa to fall back into the stampede. Simba finds and tries to wake his now-dead father before curling up under his leg, in tears. The music following was the first 4 minutes of “Adagio for Strings, Op. 11” by Samuel Barber.

The musical piece Hakuna Matata from “The Lion King” was the video for the happy MIP. The characters Timon and Pumba explain “Hakuna Matata”, which means “no worries”, to Simba, before the song ensues. Timon and Pumba then teach Simba how to eat as they do, and the clip fades out as they dance into the sunset. The music following was the first 4 minutes of “Coppélia, Act I: 1. Prélude et Mazurka,” by Léo Delibes.

While participants played the intervention game or watched the control condition video, they listened to mood inducing music from Fakhrosseini and Jeon (2017). Those in the sad condition heard “Into the Dark” by Sebastian Larsson, which is approximately 5 minutes long. Those playing Go Viral! that took over 5 minutes also listened to “At the Ivy Gate” by Brian Crain, allowing up to approximately 10 minutes to complete the game with music. The happy condition heard J. S. Bach’s “Brandenburg Concerto No.3 in G major”, which is also approximately 5 minutes long. Those playing Go Viral! that took over 5 minutes heard in addition “Brandenburg Concerto No.2 in F major” for a total of approximately 10 minutes of music.

Go Viral!

Qualtrics hosted the game which was imbedded within the survey along with an MP3 file of condition-appropriate mood music. Introductory materials described a game as using ‘the example of COVID19 misinformation to teach about techniques used in spreading online misinformation’. The page included instructions to listen to the music and play the game simultaneously, a procedure successfully trialled in the pilot study. At the bottom of the page, to check for completion of the game, were two forced-answer items asking participants their final score and the background colour on which the score was presented.

Control Video

The control condition video comprised of a sequence of screen shots of reading material from the UN Innovation Network’s Brief ‘Applying Behavioural Science to Tackle Misinformation’. The brief lists and describes state-of-the-art behavioural scientific initiatives for addressing misinformation online, but does not offer any information, training, or advice on how to protect oneself. That is, it is not itself a misinformation intervention of any sort. Along with visual stimuli was played the mood condition-congruent music detailed in the MIPs section

above. Participants were instructed to read the information presented, but to not worry too much if they don't manage to take it all in. There were two visual attention checks. Appendix C features the visual frames used in the video.

Procedure

Following recruitment to phase one, Prolific directed participants to the *Qualtrics* environment hosted by the University of Groningen. Survey materials informed prospective participants, then sought consent for their involvement and the use of the data Qualtrics would collect. Participants then gave their age and gender, and completed pre-measures on mood and misinformation susceptibility. After random selection to one of four conditions (happy/sad x Go Viral!/control), participants engaged with the MIP and intervention appropriate to their condition. They were then asked if they fully engaged with the media and were invited to leave a comment if they wished. MIPs and interventions were followed at T2 by the mood post-measure, the apprehensive and motivational threat scales, and the post-intervention measure of susceptibility to misinformation. Finally, a survey page briefed subjects, presented information related to possibilities for emotional support, and offered an opportunity to watch 'the happy video', if they wished. The final page then thanked participants and reminded them about the phase-two survey before returning them to Prolific to register completion. Participation was remunerated at £2.50 for the approximately 25-minute commitment.

The following morning, Prolific emailed invitations to participants who adequately completed phase one, inviting them to participate in a short follow-up questionnaire. This phase included briefing and consent, the T3 presentation of the MIST-20, and a more-complete debriefing on the nature of the study. Phase two took approximately 3 minutes, payment for which was £1.00.

Results

All analyses were conducted in R, version 4.2.3 (R Core Team, 2023). Assessments of data quality, the descriptive statistics, assumptions checks, and analyses of the efficacy of mood manipulations, appear in Appendix D. In Appendix E can be found a reanalysis for the efficacy of mood manipulations with a reduced sample to assess for any effect of some participants taking a long time over the survey-embedded interventions.

Hypothesis Testing

Hypothesis 1

H1 held that at T2 the inoculation condition would judge unreliable news headlines to be less reliable than would the control condition, i.e., those who played Go Viral! would have higher scores on MIST_{FAKE}. As pre-registered, this prediction was addressed via a mixed models repeated measures ANOVA with inoculation condition as a between-subjects factor and timepoint (T1 and T2) within-subjects. Results show that accuracy in rating unreliable news items was significantly improved by the advancement of timepoint, $F(1, 366) = 30.53, p < .001, \eta^2 = .01$, and that there was an interaction between condition and timepoint such that those who played Go Viral! were significantly more improved at T2, $F(1, 366) = 4.23, p = .04, \eta^2 = .001$. Thus, H1 was supported.

Hypothesis 2

H2 predicted that at T2, the happy group compared to the sad group would be a) less sceptical of all news headlines, and b) less able to discern between reliable and unreliable headlines. As pre-registered, analyses took a SDT approach, with sensitivity (d') denoting discernment and a more conservative (positive) response bias (c) indicating higher scepticism. Hit rates (H) and false alarm rates (F) for ratings of reliable and unreliable news items in both mood groups informed the measures of sensitivity, $d' = z(H) - z(F)$, and response bias, $c = -$

$0.5(z[H] + z[F])$ (Hautus et al., 2022). Correct performance on 75% of both reliable and unreliable news ratings would yield $d' = 1.35$, and 69%, $d' = 1.0$. The Gourevitch and Galanter approximation (Gourevitch & Galanter, 1967) provided variances for z -tests of statistical significance.

Discernment. The happy group had sensitivity $d' = 1.65$, and the sad group $d' = 1.34$. The difference in sensitivity at T2 was $\Delta d' = -0.31$, 99%CI [-0.38, -0.24], $p < .001$, indicating that the happy group had better discernment. Thus, H2a was not supported.

However, a pre-registered follow up analysis showed that at T1 the happy group had sensitivity $d' = 1.74$, and the sad group $d' = 1.14$, indicating the happy group had better pre-MIP discernment ($\Delta d' = -0.61$, 99%CI [-0.54, -0.67], $p < .001$). Thus, a not preregistered post-hoc difference-in-differences analysis for T1 and T2 reliability ratings between mood groups was undertaken. This showed that the sad group significantly improved discernment relative to the happy group between T1 and T2 ($\Delta d'_{\text{DIFF}} = 0.30$, 99%CI [0.20, 0.40], $p < .001$). Further, between T1 and T2, discernment significantly decreased in the happy group ($\Delta d' = -0.09$, 99%CI [-0.15, -0.03], $p < .001$), and increased in the sad group ($\Delta d' = 0.21$, 99%CI [0.13, 0.28], $p < .001$).

Similarly, for T3, preregistered exploratory analyses showed the happy group had better discernment ($d' = 1.73$ vs $d' = 1.47$, $\Delta d' = -0.26$, 99%CI [-0.31, -0.19], $p < .001$). While not preregistered, a post-hoc difference-in-differences follow up analysis showed the sad group significantly improved their discernment relative to the happy group between T1 and T3 also ($\Delta d'_{\text{DIFF}} = 0.34$, 99%CI [0.25, 0.44], $p < .001$). Figure 1 shows discernment on the MIST-20 across timepoints, by mood group.

Scepticism. A positive response bias denotes more conservative responding, which indicates scepticism. For the happy group, T2 bias was $c = 0.301$, and for the sad group $c =$

0.366, $\Delta c = 0.065$, 99%CI [0.047, 0.082], $p < .001$, indicating that the sad group were more sceptical at T2. Thus, H2b was supported.

A pre-registered follow up analysis to assess T1 differences showed that bias in the happy group was $c = 0.165$, and $c = 0.015$ in the sad group ($\Delta c = -0.150$, 99%CI [-0.184, -0.116], $p < .001$), indicating that the happy group were significantly more sceptical pre-intervention than the sad group. Because of this, a not preregistered post-hoc difference-in-differences analysis for T1 and T2 scepticism between mood groups was undertaken, and showed that the sad group became significantly more sceptical than the happy group between T1 and T2 ($\Delta c_{DIFF} = 0.215$, 99%CI [0.166, 0.263], $p < .001$). Preregistered exploratory analyses considered response bias at T3, and indicated that the sad group were more sceptical at T3 ($c = 0.311$ vs $c = 0.348$, $\Delta c = 0.037$, 99%CI [0.002, 0.071], $p < .01$). Further, a not preregistered, post-hoc difference-in-differences analysis of T1 and T3 response bias indicated that the sad group became significantly more conservative over time ($\Delta c = 0.187$, 99%CI [0.138, 0.235], $p < .001$). Figure 2 shows scepticism for MIST-20 items across timepoints, by mood group.

Hypothesis 3

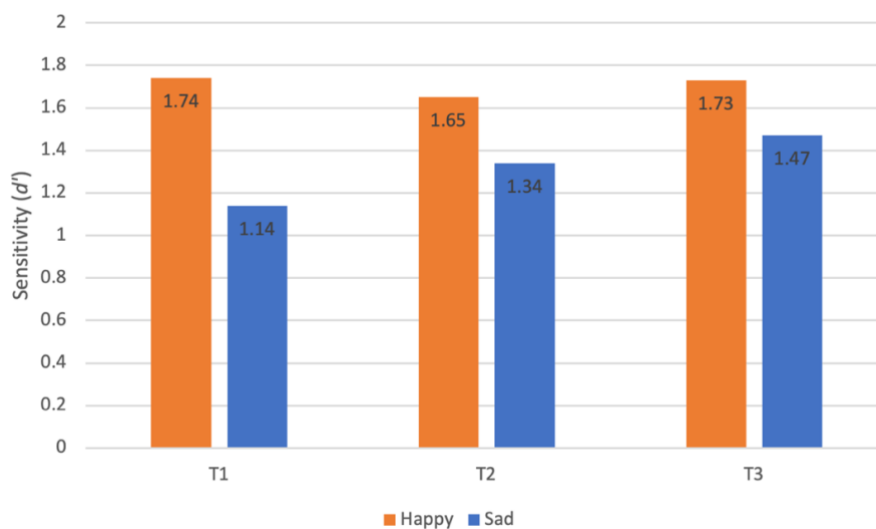
H3 predicted that, compared to the control condition, the inoculation condition would perceive more a) apprehensive threat, and b) motivational threat (T2). Robust t -tests of independent means supported H3a, $\tau(215.88) = 3.19$, $p < .001$ (one-tailed), but not H3b, $\tau(219.70) = 1.13$, $p = .13$ (one-tailed).

Hypothesis 4

H4 predicted that, compared to the happy group, the sad group would perceive more a) apprehensive threat, and b) motivational threat (T2). Robust t -tests of independent means supported H4a, $\tau(98.49) = 1.76$, $p = .04$ (one-tailed), but not H4b, $\tau(108.50) = 0.24$, $p = .40$ (one-tailed).

Figure 1

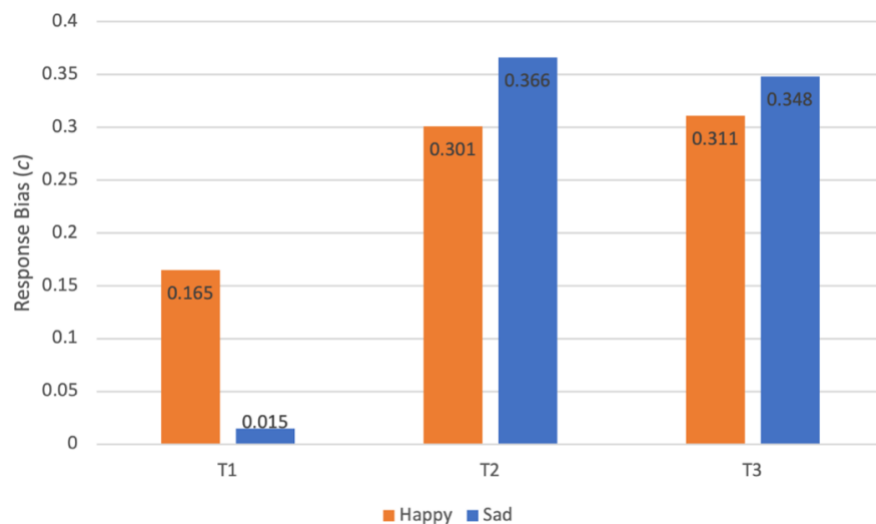
Discernment on the MIST-20 Across Timepoints, by Mood Group



Note. Differences between groups at each timepoint were significant for $\alpha = .001$, as were T1-T2 differences for each group, and the difference-in-differences T1-T2, and T1-T3.

Figure 2

Scepticism for MIST-20 Items Across Timepoints, by Mood Group



Note. Differences between groups at each timepoint were significant for $\alpha = .001$, as were T1-T2 differences for each group, and the difference-in-differences T1-T2, and T1-T3.

Hypothesis 5

H5 held that at T3 the inoculation condition, compared to the control condition, would judge unreliable news headlines to be less reliable. As pre-registered, this prediction was addressed via a mixed models repeated measures factorial ANOVA with inoculation condition as a between-subjects factor and timepoint (T1 and T3) a within-subjects factor. Results show that accuracy in recognising unreliable news items was significantly improved by the advancement of test timepoint, $F(1, 348) = 32.03, p < .001, \eta^2 = .01$, and an interaction between condition and timepoint such that those who played Go Viral! were significantly more improved at T3, $F(1, 348) = 4.23, p = .02$ (one-tailed), $\eta^2 = .001$. Thus, H5 was supported.

Hypothesis 6

H6 held that for those that played Go Viral!, those in the sad group would judge unreliable news headlines to be less reliable at T3 than those in the happy group. As pre-registered, this prediction was addressed via a mixed models repeated measures ANOVA with inoculation condition as a between-subjects factor and timepoint (T1 and T3) within-subjects. Results show that accuracy in recognising unreliable news items was significantly improved by the advancement of test timepoint, $F(1, 169) = 27.49, p < .001, \eta^2 = .02$, but that the interaction between condition and timepoint was not associated with a significant improvement in T3 ratings of unreliable news items, $F(1, 169) = 0.07, p = 0.39$ (one-tailed). Thus, H6 was not supported.

Effects of Go Viral!

As pre-registered, assessments were made of differences between inoculation and control condition discernment (sensitivity, d') and scepticism (response bias, c), at T2 and T3. It was expected that there would be no difference in discernment, but that the inoculation condition would be more sceptical, i.e., tend more toward 'unreliable news' responses.

Discernment. At T2, the Go Viral! condition had sensitivity $d' = 1.45$ and the control condition $d' = 1.57$, $\Delta d' = -0.12$. The confidence interval from an equivalence test with the threshold for equivalence set at ± 0.5 standard deviations was $CI[-0.13, -0.11]$, indicating that discernment between groups was not equivalent, violating expectations of no difference. A preregistered follow-up significance test indicated that the control group had better T2 discernment, $\Delta d' = -0.12$, 99%CI $[-0.19, -0.05]$, $p < .001$.

At T3 the Go Viral! condition had sensitivity $d' = 1.52$ and the control condition $d' = 1.62$, $\Delta d' = -0.10$. The confidence interval from an equivalence test with the threshold for equivalence set at ± 0.5 standard deviations was $CI[-0.12, -0.09]$, indicating that discernment between groups was not equivalent, violating expectations of no difference. The preregistered follow-up significance test indicated that the control group had better T3 discernment, $\Delta d' = -0.10$, 99%CI $[-0.17, -0.04]$, $p < .001$.

Though not preregistered, a post-hoc exploratory analysis of differences in discernment between the Go Viral! and control groups at T1 showed that sensitivity was identical ($d' = 1.29$). Figure 3 shows discernment on the MIST-20 across timepoints, by intervention condition.

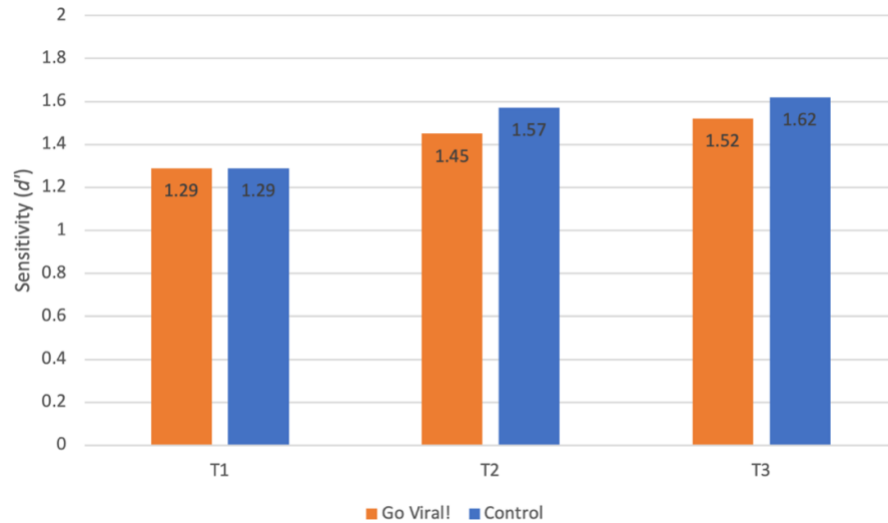
Scepticism. At T2, the Go Viral! condition had response bias $c = 0.312$ and the control condition $c = 0.343$, $\Delta c = -0.031$, 99%CI $[-0.066, 0.004]$, NS for $\alpha = .01$. Results thus indicated no significant difference in scepticism by intervention condition, violating an expectation the Go Viral! group would be more sceptical. At T3 the Go Viral! condition had response bias $c = 0.320$ and the control condition $c = 0.315$, $\Delta c = 0.005$, 99%CI $[-0.29, 0.039]$, NS for $\alpha = .01$, again indicating no significant difference in scepticism, against expectations.

Though not preregistered, a post-hoc exploratory analysis of differences in response bias between the Go Viral! and control groups at T1 showed response bias was identical ($c = 0$).

Figure 4 shows scepticism on the MIST-20 across timepoints, by intervention condition.

Figure 3

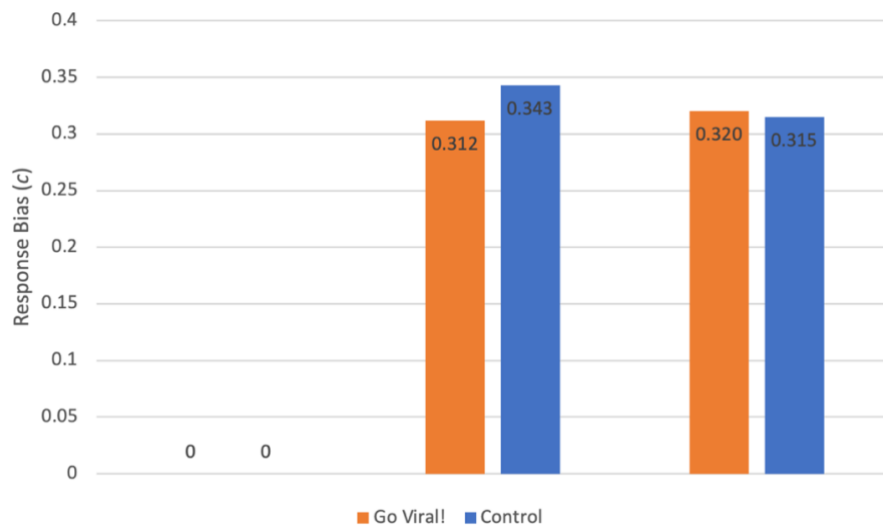
Discernment on the MIST-20 Across Timepoints, by Intervention Condition



Note. Differences between groups at T2 and T3 were significant for $\alpha = .001$.

Figure 4

Scepticism on the MIST-20 Across Timepoints, by Intervention Condition



Note. No within timepoint differences between conditions were significant for $\alpha = .01$.

Mediation and Conditional Process Analyses

Pre-registered analyses were conducted to assess any mediating role of perceived threat (apprehensive and motivational) between inoculation condition and ratings of unreliable news headlines at T3, and any moderating role of mood on the inoculation/perceived threat pathway. However, correlations between each type of perceived threat and MIST_{FAKE} ratings at T3 were nonsignificant ($r = .14$ for both types of threat and MIST_{FAKE} T3 in the full sample, and between $r = .10$ and $.14$ by inoculation group. There was power to detect $r = .15$). Thus, mediation, and by extension, moderated mediation, were not possible outcomes and any further analysis was abandoned.

Discussion

The primary questions the current study sought to address concerned the influence of mood and playing inoculation game Go Viral! on susceptibility to misinformation. They also queried into any differences there may be in two types of perceived threat, depending on mood and playing the game. Secondary questions related specifically to changes in discernment and scepticism for reliable and unreliable news headlines from playing Go Viral!, and to the mediating role of perceived threat on susceptibility to misinformation, along with any moderating role there may be of mood.

Effects of Mood on Susceptibility to Misinformation

Results showed that the happy group were more rather than less discerning at T2, but less sceptical than the sad group. However, follow-up analyses showed that there was significant and substantial T1 heterogeneity in mood groups on both measures, with initial discernment and scepticism higher in the happy group. The full pattern of results clearly shows that post MIPs, discernment was significantly boosted in the sad group and depleted in the happy group, while both groups became more sceptical, but the increase was significantly more for the sad group.

Thus, the sad group realised greater gains in discernment and scepticism after mood inductions, with mood groups matched on intervention conditions.

Outcomes were consistent with a prediction of the AIM that a sad mood will result in more-accurate discernment and more scepticism for reliable and unreliable information than a happy mood. By this account, mood influences information processing such that a sad mood will promote an accommodative processing style associated with more elaboration and thus better discernment, while a happy mood will promote an assimilative processing style associated with a more heuristic orientation to information, and thus poorer discernment. Further, each mood will be associated with informational effects, such as those of affect priming and affect-as-information, that will drive mood-congruent judgements of information such that happier people will be more gullible and sadder people more sceptical. The fact that participants in a happier mood became more sceptical can be attributed in this case to the intervention and control conditions, the MIST-20 susceptibility test, and tests for perceived threat, all making salient the topic of online misinformation. This would likely have promoted conservative vigilance in participants. The retention in both mood groups of T2 scepticism at T3, the next day, may be due to the topic of misinformation being re-invoked by test materials so soon after the initial survey and interventions, and thus becoming salient once more.

At T3 the sad group had improved discernment again, as had the happy group, although in the latter case it was merely back to T1 levels. T3 results for discernment in the sad group can be explained by sadder individuals attending more to stimuli presented before and during T2, that is, the intervention and susceptibility test materials, and thus noticing judgment cues in the test items at T3 that allowed further improvements in rating accuracy. Alternately, happier individuals were less attentive at T2 and simply regained their original discernment abilities at T3, once MIPs had worn off.

Perceived Threat from Mood and Go Viral!

An important consideration when testing predictions of the AIM for effects of mood on judgements, such as for those concerning online (mis)information, is the presence of any motivating pressure. This is because any such prediction would include that a motivator would attenuate or eliminate effects of mood on such judgements. Perceived threat is equally important to address in the context of an inoculation intervention such as Go Viral!, as the theoretical grounding of the game holds that threat provides a necessary motivational drive.

Results showed that Go Viral! and sadness were associated with higher apprehensive threat, the traditional measure of threat used in inoculation research, but that there were no differences in motivational threat, a newer measure of threat more closely related to the theorised processes of inoculation. This is the opposite of what was found in Basol et al. (Study 1; 2021), the only study into an inoculation game to consider perceived threat.

Also planned was an analysis of the mediating roles of each kind of perceived threat, between playing Go Viral! and T3 ratings of unreliable news headlines. Pearson correlations showed there were no associations between variables in this sample. Thus, neither apprehensive nor motivational perceived threat moderated the pathway between playing Go Viral! and accurate ratings of unreliable news headlines at T3, against an expectation that, being an inoculation intervention, they would.

Effects of Playing Go Viral!

The effect of playing Go Viral! on T2 ratings of unreliable news headlines compared to browsing non-instructive, general information about misinformation interventions via video, with mood balanced within conditions, served as a check on the effect of the game on susceptibility to misinformation by the method that has been most commonly employed in studies to date. Results were consistent with past studies into inoculation games, all of which

have demonstrated that unreliable, or ‘fake’ news items, are judged as less reliable (or less likely true, or more manipulative etc.), after playing an inoculation game (Basol et al., 2020, 2021; Maertens, Götz, et al., 2021; Maertens, Roozenbeek, et al., 2021; Roozenbeek & van der Linden, 2019a, 2020). The effect of playing Go Viral! on ratings of unreliable news headlines at T3, compared to the control activity, was effectively identical to that at T2. For there to be no change in inoculation outcomes after one day of an intervention is consistent with previous longitudinal studies of inoculation games in which effects have been retained after at least a week (Basol et al., 2021; Maertens, Roozenbeek, et al., 2021).

Results of the effect of playing Go Viral! on susceptibility to misinformation considered in terms of discrimination and scepticism for reliable and unreliable news headlines, though, inspires a quite different interpretation, although still consistent with past such analyses of inoculation games (Maertens, Roozenbeek, et al., 2021; Modirrousta-Galian & Higham, 2022). Regarding discernment, at each successive timepoint the control group achieved better accuracy than the treatment group. For scepticism, there was no significant difference between groups, although both those who played Go Viral! and those who watched the control video were more sceptical at T2 and T3 than at T1. This finding was contrary to an expectation that the inoculation group would be more sceptical than the control group. While the materials of the control condition did not deliver a warning any more powerful than to refer to misinformation as ‘a growing global issue’, which it did on just one occasion, one interpretation of this result is that it may be that simply making the topic salient is enough to drive scepticism to the same extent that Go Viral! and Bad News appear to.

Effect of Mood on Inoculation

At T3, a day after the main survey and interventions, current mood state was disentangled from the mood state induced prior to and during Go Viral! gameplay. Thus, findings at T3 allow for inferences regarding the effect of mood at the time of inoculation, independent of mood at time of testing. In applying the AIM, the prediction was that sad mood at T2 would be associated with a stronger inoculation effect of the game, and therefore less susceptibility to misinformation. This would be expected due to informational and processing effects related to mood. However, there was no effect of mood found for T3 ratings of the unreliable news items of the MIST-20, for those who played Go Viral!.

Implications of the Current Findings

Overall, findings suggest there is indeed an effect of mood on susceptibility to online misinformation. Just as Forgas and colleagues found in their studies into mood and veracity judgements, effects of discernment and scepticism for reliable and unreliable news headlines were such that happiness was associated with lower levels of both. To the best of our knowledge, this is the first study to demonstrate this effect experimentally using a scale designed to emulate online misinformation. This extends the literature on mood and judgement, which has focused on interpersonal settings rather than information embedded in a news- or social media-style format. A serious implication of the generalisation of this effect to include online misinformation concerns the practice of temporal micro-targeting of persuasive messages by mood, via which manipulative communications may be strategically presented while the consumer is most receptive (Dawson, 2020; Lewandowsky & van der Linden, 2021). Mood states have been accurately inferred from user generated data by various computational means, including natural language processing techniques such as textual emotion detection (Nimeshika & Ahangama, 2019; Saffar et al., 2023). If such patterns exist and are indeed detectable by machine learning

algorithms, then they may be exploited by the same to maximise the impact of whatever information they are tasked with promoting, which could include misinformation and political propaganda. Conversely, the same patterns could inform as to when a person is most at need of support in resisting online misinforming attacks.

A further implication of the current study stems from the lack of evidence here for an effect of mood on the process of inoculation to misinformation that Go Viral! is designed to initiate. Moreover, motivational perceived threat was not higher in those that played Go Viral! compared to a non-inoculating control condition, and nor was there any mediating role of threat in intervention on outcomes, as theory would predict. The discussion following presents two possibilities for these findings.

The first possibility is that mood does not play a role in inoculation, and neither does motivational threat. The implications of this for inoculation theory are consequential in that the removal of motivational threat as a mechanism of inoculation would remove also the need for an initial warning step, leaving only a prebunk. That is to say, in this case at least, prebunking and inoculation would be one-in-the-same. In the original theory and classic experiments, forewarning was tied to a theoretical assumption that inoculation can only occur in the ‘germ free’ environment of cultural truisms – beliefs for which people would be unaware there exist counterarguments (McGuire, 1964). Thus, in lieu of any natural motivation to protect a belief that is assumed to be beyond the possibility of attack, motivation to engage must first be instilled. In the context of contested issues, though, people are well aware that there exist counter positions to their own, thus rendering the original rationale for a warning step baseless. Indeed, earlier findings that inoculation on controversial topics did not lead to the production of counterarguments, and that refutations conferred no more resistance to attitude change than supportive arguments, have led to the suggestion that inoculation on contested topics may not be

inoculation proper, just as a prebunk is not inoculation by itself (see Benoit, 1991 for a discussion on inoculation on controversial topics, and Traberg et al, 2022 for a discussion on inoculation as more than prebunking, by definition).

The second possibility is that Go Viral! does not inoculate against susceptibility to misinformation. This is a possibility one might entertain in heeding inoculation researchers Compton and Pfau (2005) who wrote, ‘simply stated, inoculation is impossible without threat’ (pp. 100-101). Indeed, to apply here motivational threat as a manipulation check for inoculation in the same way the traditional, less theoretically relevant measure of threat was used for decades, would be to determine the manipulation was unsuccessful. Relatedly, it may be argued that discernment, rather than ratings of unreliable news headlines only, better captures both a specific effect of resistance to misinformation and a more adaptive outcome (Guay et al., 2022). If one does prefer discernment over ratings of unreliable news items only as a measure of resistance to persuasion from misinformation, then one would find evidence in the current findings that reading general descriptions of methods to address online misinformation, yet without experiencing anything of those interventions, is effective for conferring protection to online misinformation, but playing the inoculation game Go Viral! is not. If this is the case, it follows then that playing any inoculation game based on Bad News, for example Breaking Harmony Square, Radicalize, and Under Pressure, may not confer resistance to misinformation.

Limitations and Future Directions

The current study has several limitations that future research might address. Firstly, if either possibility outlined above are the case, they being that inoculation in the context of contested issues is not inoculation *per se*, or that the inoculation game Go Viral! does not inoculate, then the current study has the serious limitation of not being equipped to make inferences regarding its primary focus – the role of mood in psychological inoculation. In either

case, the theoretical argument outlined here remains in contention and might best be explored within a more traditional inoculation paradigm.

Also, happy and sad mood states considered here are not exemplars of positive and negative mood states generally, and future research into mood, inoculation, and susceptibility to misinformation might include others. For example, in contrast to sadness, the negative mood state anger may increase susceptibility to misinformation and thus also have different effects on inoculation than sadness (Miller et al., 2013; Sharma et al., 2023).

Further, if inoculation games do not inoculate against misinformation specifically, a fuller understanding of the consequences of what they do achieve, and in different populations, would be pertinent. While scepticism for all information may well be adaptive in protecting against persuasion via techniques that do not tend to have identifiable characteristics, like ‘deep fake’ and de-contextualised video footage, boosting scepticism in people that have already high levels of distrust in news media could reach tipping points into the extremism associated with conspiratorial ideation (Hayward & Gronland, 2021; van Prooijen & Douglas, 2018).

Specifically in relation to Go Viral!, the invocation of apprehensive and motivational threat from playing the game is still unclear as the current study found opposite effects to those of Basol et al. (2021). We recommend future research clarify those relationships.

A further limitation that future research might seek to address is that it took twice as long for participants to play Go Viral!, marketed as a 5-minute game, than to watch a 5-minute video. Why this was is not clear, nor whether this trend is indeed related to the relatively less effective MIPs for those who played the game. It could be that online participants, left to their own devices, were not engaged with the game, which could explain why it didn’t seem to work (by the metric of discernment, at least). However, a very high proportion of players reported they did engage, and all passed checks to confirm they completed the game. It may be that the game did

not drive interest, but if that were so it would represent a flaw of the intervention on a par with other potential flaws. In future, researchers might match the duration of the control condition to a more accurate estimate of Go Viral! gametime, and conduct experiments in-person to broach this potential confound.

Finally, although social desirability was not implicated in mood ratings in the pilot study, Banas & Richards (2017) point out that the measures of perceived threat in the main study are transparent and may be subject to such bias. It is recommended that future studies thus incorporate a measure of social desirability so this potential influence may be controlled for.

Conclusion

While the present study fell short of its primary goal of elaborating the roles of mood and threat in inoculation, taking an inoculation game as the treatment intervention, results integrated into past research and considered within the frameworks of Psychological Inoculation Theory and the Affect Infusion Model are suggestive of a meaningful contribution to be followed up in future research. Arguably, the most pressing question of those presented here pertains to the metric by which researchers might gauge the success or otherwise of interventions to reduce susceptibility to misinformation. With results offering opposite interpretations by two methods that are both currently accepted for publication, some consensus as to which is preferred must surely come before determinations of efficacy can be plausibly made.

References

- Banas, J. A., & Richards, A. S. (2017). Apprehension or motivation to defend attitudes? Exploring the underlying threat mechanism in inoculation-induced resistance to persuasion. *Communication Monographs*, *84*(2), 164–178.
<https://doi.org/10.1080/03637751.2017.1307999>
- Basol, M., Roozenbeek, J., Berriche, M., Uenal, F., McClanahan, W. P., & Linden, S. van der. (2021). Towards psychological herd immunity: Cross-cultural evidence for two prebunking interventions against COVID-19 misinformation. *Big Data & Society*, *8*(1), 205395172110138. <https://doi.org/10.1177/20539517211013868>
- Basol, M., Roozenbeek, J., & Van der Linden, S. (2020). Good news about Bad News: Gamified inoculation boosts confidence and cognitive immunity against fake news. *Journal of Cognition*, *3*(1), article 2. <https://doi.org/10.5334/joc.91>
- Benoit, W. L. (1991). Two tests of the mechanism of inoculation theory. *Southern Communication Journal*, *56*(3), 219–229. <https://doi.org/10.1080/10417949109372832>
- Bless, H., & Fiedler, K. (2006). Mood and the regulation of information processing and behavior. In J. P. Forgas (Ed.), *Affect in social thinking and behavior* (pp. 65–84). Psychology Press.
- Bower, G. H. (1981). Mood and memory. *American Psychologist*, *36*, 129–148.
- Bruder, M., Haffke, P., Neave, N., Nouripanah, N., & Imhoff, R. (2013). Measuring individual differences in generic beliefs in conspiracy theories across cultures: Conspiracy Mentality Questionnaire. *Frontiers in Psychology*, *4*. <https://doi.org/10.3389/fpsyg.2013.00225>
- Burgoon, M., Miller, M. D., Cohen, M., & Montgomery, C. L. (1978). An empirical test of a model of resistance to persuasion. *Human Communication Research*, *5*(1), 27–39.
<https://doi.org/10.1111/j.1468-2958.1978.tb00620.x>

- Chan, M. S., Jones, C. R., Hall Jamieson, K., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science*, 28(11), 1531–1546. <https://doi.org/10.1177/0956797617714579>
- Clark, L. A., & Watson, D. (1994). *The PANAS-X: Manual for the Positive and Negative Affect Schedule - Expanded Form* [Data set]. <https://doi.org/10.17077/48vt-m4t2>
- Compton, J., Ivanov, B., & Hester, E. (2022). Inoculation Theory and Affect. *International Journal of Communication*, 16, 3470-3483.
- Compton, J., & Pfau, M. (2005). Inoculation theory of resistance to influence at maturity: Recent progress in theory development and application and suggestions for future research. *Annals of the International Communication Association*, 29(1), 97–146. <https://doi.org/10.1080/23808985.2005.11679045>
- Cook, J., Lewandowsky, S., & Ecker, U. K. H. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PLOS ONE*. <https://doi.org/10.1371/journal.pone.0175799>
- Cook, J., Oreskes, N., Doran, P. T., Anderegg, W. R. L., Verheggen, B., Maibach, E. W., Carlton, J. S., Lewandowsky, S., Skuce, A. G., Green, S. A., Nuccitelli, D., Jacobs, P., Richardson, M., Winkler, B., Painting, R., & Rice, K. (2016). Consensus on consensus: A synthesis of consensus estimates on human-caused global warming. *Environmental Research Letters*, 11(4), 048002. <https://doi.org/10.1088/1748-9326/11/4/048002>
- Dawson, J. (2020). Microtargeting as Information Warfare [Preprint]. *SocArXiv*. <https://doi.org/10.31235/osf.io/5wzuq>
- Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The truth about the truth: A meta-analytic review of the truth effect. *Personality and Social Psychology Review*, 14(2), 238–257. <https://doi.org/10.1177/1088868309352251>

- Donnelly, S., Jorgensen, T.D. & Rudolph, C.W. (2022). Power analysis for conditional indirect effects: A tutorial for conducting Monte Carlo simulations with categorical exogenous variables. *Behavioural Research*. <https://doi.org/10.3758/s13428-022-01996-0>
- Fakhrhosseini, S. M., & Joen, M. (2017). Affect/emotion induction methods. In M. Jeon (Ed.), *Emotions and affect in human factors and human-computer interaction* (pp. 235-253). Academic Press.
- Fanselow, M. S. (2018). Emotion, motivation and function. *Current Opinion in Behavioral Sciences*, 19, 105–109. <https://doi.org/10.1016/j.cobeha.2017.12.013>
- Forgas, J. P. (1995). Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin*, 117(1), 39–66. <https://doi.org/10.1037/0033-2909.117.1.39>
- Forgas, J. P. (2002). Feeling and doing: Affective influences on interpersonal behavior. *Psychological Inquiry*, 13(1), 1–28. https://doi.org/10.1207/S15327965PLI1301_01
- Forgas, J. P. (2007). When sad is better than happy: Negative affect can improve the quality and effectiveness of persuasive messages and social influence strategies. *Journal of Experimental Social Psychology*, 43(4), 513–528. <https://doi.org/10.1016/j.jesp.2006.05.006>
- Forgas, J. P. (2013). Don't worry, be sad! On the cognitive, motivational, and interpersonal benefits of negative mood. *Current Directions in Psychological Science*, 22(3), 225–232. <https://doi.org/10.1177/0963721412474458>
- Forgas, J. P. (2019). Happy believers and sad skeptics? Affective influences on gullibility. *Current Directions in Psychological Science*, 28(3), 306–313. <https://doi.org/10.1177/0963721419834543>

- Forgas, J. P., & East, R. (2008). On being happy and gullible: Mood effects on skepticism and the detection of deception. *Journal of Experimental Social Psychology, 44*(5), 1362–1367. <https://doi.org/10.1016/j.jesp.2008.04.010>
- Godlee, F., Smith, J., & Marcovitch, H. (2011). Wakefield's article linking MMR vaccine and autism was fraudulent: Clear evidence of falsification of data should now close the door on this damaging vaccine scare. *BMJ: British Medical Journal, 342*, 64–66. <https://doi.org/10.1136/bmj.c7452>
- Gourevitch, V., & Galanter, E. (1967). A significance test for one-parameter isosensitivity functions. *Psychometrika, 32*(1), 25–33. <https://doi.org/10.1007/BF02289402>
- Guay, B., Berinsky, A., Pennycook, G., & Rand, D. G. (2022). How to think about whether misinformation interventions work [Preprint]. *PsyArXiv*. <https://doi.org/10.31234/osf.io/gv8qx>
- Habgood-Coote, J. (2019). Stop talking about fake news! *Inquiry, 62*(9–10), 1033–1065. <https://doi.org/10.1080/0020174X.2018.1508363>
- Hayward, J., & Gronland, G. (2021). *Conspiracy Theories in the Classroom Guidance for teachers*. Arts and Humanities Research Council (UK).
- Higham, P. A., Zawadzka, K., & Hanczakowski, M. (2016). Internal mapping and its impact on measures of absolute and relative metacognitive accuracy. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 39–61). Oxford University Press. <http://dx.doi.org/10.1093/oxfordhb/9780199336746.013.15>
- Hautus, M. J., Macmillan, N. A., & Creelman, C. D. (2021). *Detection theory: A user's guide*. Routledge.
- Ivanov, B., Hester, E. B., Martin, J. C., Silberman, W., Slone, A. R., Goatley-Soan, S., Geegan, S., Parker, K. A., Herrington, T. F., Riker, S., & Anderson, A. (2020). Persistence of

- emotion in the process of inoculation: Experiencing post-attack threat, fear, anger, happiness, sadness, and surprise. *Communication Quarterly*, 68(5), 560–582.
<https://doi.org/10.1080/01463373.2020.1850492>
- Iyengar, S., & Massey, D. S. (2019). Scientific communication in a post-truth society. *Proceedings of the National Academy of Sciences*, 116(16), 7656–7661.
<https://doi.org/10.1073/pnas.1805868115>
- Jolley, D., & Paterson, J. L. (2020). Pylons ablaze: Examining the role of 5G COVID-19 conspiracy beliefs and support for violence. *British Journal of Social Psychology*, 59(3), 628–640. <https://doi.org/10.1111/bjso.12394>
- Kata, A. (2010). A postmodern Pandora's box: Anti-vaccination misinformation on the Internet. *Vaccine*, 28(7), 1709–1716. <https://doi.org/10.1016/j.vaccine.2009.12.022>
- Kroenke, K., Spitzer, R. L., Williams, J. B. W., & Löwe, B. (2010). The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: A systematic review. *General Hospital Psychiatry*, 32(4), 345–359.
<https://doi.org/10.1016/j.genhosppsy.2010.03.006>
- Lewandowsky, S., Cook, J., & Ecker, U. K. H. (2017). Letting the gorilla emerge from the mist: Getting past post-truth. *Journal of Applied Research in Memory and Cognition*, 6(4), 418–424. <https://doi.org/10.1016/j.jarmac.2017.11.002>
- Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13(3), 106–131.
<https://doi.org/10.1177/1529100612451018>

- Lewandowsky, S., Lloyd, E. A., & Brophy, S. (2017). When THUNCIing trumps thinking: What distant alternative worlds can tell us about the real world. *Argumenta*, 3(2), 1–15.
<https://doi.org/10.23811/52.arg2017.lew.llo.bro>
- 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>
- Loomba, S., de Figueiredo, A., Piatek, S. J., de Graaf, K., & Larson, H. J. (2021). Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nature Human Behaviour*, 5(3), 337–348. <https://doi.org/10.1038/s41562-021-01056-1>
- Maertens, R., Götz, F. M., Golino, H., Roozenbeek, J., Schneider, C. R., Kyrychenko, Y., Kerr, J. R., Stieger, S., McClanahan, W. P., Drabot, K., He, J. K., & Linden, S. van der. (2021). The Misinformation Susceptibility Test (MIST): A psychometrically validated measure of news veracity discernment [Preprint]. *PsyArXiv*. <https://doi.org/10.31234/osf.io/gk68h>
- Maertens, R., Roozenbeek, J., Basol, M., & van der Linden, S. (2021). Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments. *Journal of Experimental Psychology: Applied*, 27(1), 1–16. <https://doi.org/10.1037/xap0000315>
- Marcusson-Clavertz, D., Kjell, O. N. E., Persson, S. D., & Cardeña, E. (2019). Online validation of combined mood induction procedures. *PLOS ONE*, 14(6), e0217848.
<https://doi.org/10.1371/journal.pone.0217848>
- McCornack, S. A., & Parks, M. R. (1986). Deception detection and relationship development: The other side of trust. *Annals of the International Communication Association*, 9(1), 377–389. <https://doi.org/10.1080/23808985.1986.11678616>

- McCright, A. M., Charters, M., Dentzman, K., & Dietz, T. (2016). Examining the effectiveness of climate change frames in the face of a climate change denial counter-frame. *Topics in Cognitive Science*, 8(1), 76–97. <https://doi.org/10.1111/tops.12171>
- McGuire, W. J. (1964). Some contemporary approaches. In *Advances in Experimental Social Psychology* (Vol. 1, pp. 191–229). Elsevier. [https://doi.org/10.1016/S0065-2601\(08\)60052-0](https://doi.org/10.1016/S0065-2601(08)60052-0)
- McGuire, W. J. (1970). Vaccine for brainwash. *Psychology today*, 3(9), 36–64.
- McIntyre, L. (2018). *Post-truth*. MIT Press.
- Miller, C. H., Ivanov, B., Sims, J. D., Compton, J., Harrison, K. J., Parker, K. A., Parker, J. L., & Averbek, J. M. (2013). Boosting the potency of resistance: Combining the motivational forces of inoculation and psychological reactance. *Human Communication Research*, 39(1), 127–155. <https://doi.org/10.1111/j.1468-2958.2012.01438.x>
- Modirrousta-Galian, A., & Higham, P. A. (2022). How Effective are Gamified Fake News Interventions? Reanalyzing Existing Research with Signal Detection Theory. *PsyArxiv Preprints*. <https://www.doi.org/10.31234/osf.io/4bgkd>
- Nimeshika, S., & Ahangama, S. (2019). A method to identify the current mood of social media users. *2019 14th Conference on Industrial and Information Systems (ICIIS)*, 356–359. <https://doi.org/10.1109/ICIIS47346.2019.9063291>
- Pennycook, G., Allan Cheyne, J., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2015). On the reception and detection of pseudo-profound bullshit. *Judgment and Decision Making*, 10(6), 549–563. <https://doi.org/10.1017/S1930297500006999>
- Pfau, M. (1997). The inoculation model of resistance to influence. In G. A. Barnett & F. J. Boster (Eds.), *Progress in communication sciences: Advances in persuasion* (Vol. 13, pp. 133–171). Ablex.

- Pfau, M., Semmler, S. M., Deatrick, L., Mason, A., Nisbett, G., Lane, L. T., Craig, E. A., Jill Underhill, Jill Underhill, Underhill, J. C., & Banas, J. A. (2009). Nuances about the role and impact of affect in inoculation. *Communication Monographs*, 76(1), 73–98.
<https://doi.org/10.1080/03637750802378807>
- Pfau, M., Szabo, A., Anderson, J., Morrill, J., Zubric, J., & H-Wan, H.-H. (2001). The role and impact of affect in the process of resistance to persuasion. *Human Communication Research*, 27(2), 216–252. <https://doi.org/10.1111/j.1468-2958.2001.tb00781.x>
- Pine, R., Fleming, T., McCallum, S., & Sutcliffe, K. (2020). The effects of casual videogames on anxiety, depression, stress, and low mood: A systematic review. *Games for Health Journal*, 9(4), 255–264. <https://doi.org/10.1089/g4h.2019.0132>
- R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Revelle, W., & Condon, D. M. (2019). Reliability From α to ω : A Tutorial. *Psychological Assessment*, 31(12), 1395-1411.
- Richards, A. S., & Banas, J. A. (2018). The opposing mediational effects of apprehensive threat and motivational threat when inoculating against reactance to health promotion. *Southern Communication Journal*, 83(4), 245–255.
<https://doi.org/10.1080/1041794X.2018.1498909>
- Roozenbeek, J., Maertens, R., Herzog, S. M., Geers, M., Kurvers, R., Sultan, M., & van der Linden, S. (2022). Susceptibility to misinformation is consistent across question framings and response modes and better explained by myside bias and partisanship than analytical thinking. *Judgment and Decision Making*, 17(3), 547–573.
<https://doi.org/10.1017/S1930297500003570>

- Roozenbeek, J., Maertens, R., McClanahan, W., & Van Der Linden, S. (2021). Disentangling Item and Testing Effects in Inoculation Research on Online Misinformation: Solomon Revisited. *Educational and Psychological Measurement, 81*(2), 340–362.
<https://doi.org/10.1177/0013164420940378>
- Roozenbeek, J., & van der Linden, S. (2019a). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications, 5*(1), 65.
<https://doi.org/10.1057/s41599-019-0279-9>
- Roozenbeek, J., & van der Linden, S. (2019b). The fake news game: Actively inoculating against the risk of misinformation. *Journal of Risk Research, 22*(5), 570–580.
<https://doi.org/10.1080/13669877.2018.1443491>
- Roozenbeek, J., & van der Linden, S. (2020). Breaking Harmony Square: A game that “inoculates” against political misinformation. *Harvard Kennedy School Misinformation Review*. <https://doi.org/10.37016/mr-2020-47>
- Rupp, M. A., Sweetman, R., Sosa, A. E., Smither, J. A., & McConnell, D. S. (2017). Searching for affective and cognitive restoration: Examining the restorative effects of casual video game play. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 59*(7), 1096–1107. <https://doi.org/10.1177/0018720817715360>
- Saffar, A. H., Mann, T. K., & Ofoghi, B. (2023). Textual emotion detection in health: Advances and applications. *Journal of Biomedical Informatics, 137*, 104258.
<https://doi.org/10.1016/j.jbi.2022.104258>
- Schwarz, N. (1990). Feelings as information: Informational and motivational functions of affective states. In E. T. Higgins & R. Sorrentino (Eds.), *Handbook of motivation and cognition* (Vol. 2, pp. 527–561). Guilford Press.
- Steyer, R., Schwenkmezger, P., Notz, P. und Eid, M. (1997). Der Mehrdimensionale

Befindlichkeitsfragebogen (MDBF). Hogrefe.

Sharma, P. R., Wade, K. A., & Jobson, L. (2023). A systematic review of the relationship between emotion and susceptibility to misinformation. *Memory*, *31*(1), 1–21.

<https://doi.org/10.1080/09658211.2022.2120623>

Stöber, J. (2001). The Social Desirability Scale-17 (SDS-17). *European Journal of Psychological Assessment*, *17*(3), 222–232. <https://doi.org/10.1027//1015-5759.17.3.222>

Traberg, C. S., Roozenbeek, J., & van der Linden, S. (2022). Psychological inoculation against misinformation: Current evidence and future directions. *The ANNALS of the American Academy of Political and Social Science*, *700*(1), 136–151.

<https://doi.org/10.1177/00027162221087936>

van der Linden, S., & Roozenbeek, J. (2020). Psychological inoculation against fake news. In R. Greifeneder, M. Jaffé, E. J. Newman, & N. Schwarz (Eds.), *The psychology of fake news: Accepting, sharing, and correcting misinformation* (pp. 147–169). Psychology Press.

van Prooijen, J.-W., & Douglas, K. M. (2018). Belief in conspiracy theories: Basic principles of an emerging research domain. *European Journal of Social Psychology*, *48*(7), 897–908.

<https://doi.org/10.1002/ejsp.2530>

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>

Walter, N., & Tukachinsky, R. (2020). A meta-analytic examination of the continued influence of misinformation in the face of correction: How powerful is it, why does it happen, and how to stop it? *Communication Research*, *47*(2), 155–177.

<https://doi.org/10.1177/0093650219854600>

Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing* (2nd ed.).

Elsevier.

World health Organization. (2020). Munich Security Conference speech. www.who.int.

Wu, L., Morstatter, F., Carley, K. M., & Liu, H. (2019). Misinformation in social media:

Definition, manipulation, and detection. *ACM SIGKDD Explorations Newsletter*, 21(2),

80–90. <https://doi.org/10.1145/3373464.3373475>

Index of Appendices

Appendix A: Pilot Study.....	51
Appendix B: Scales and Adaptations.....	68
Appendix C: Visual Stimuli from the Control Condition Video.....	72
Appendix D: Main Study Descriptive Statistics, Assumption Testing, and Mood Manipulation Checks.....	75
Appendix E: Mood Manipulation Checks with Reduced Sample.....	86

Appendix A

Pilot Study

Introduction

This preregistered study aimed to address preliminary questions regarding main effects of an inoculation game and mood on susceptibility to misinformation. *Bad News* was chosen as the inoculation game with the most established positive effects on misinformation susceptibility, and the moods were happy, sad, and neutral. Effects of mood on gullibility for claims that specifically resemble online misinformation have not been demonstrated in previous research, so a preliminary enquiry into the appropriateness of these items, and the MIST-8 scale that is preferred for its brevity, is required. Findings will inform study design decision for the main study, to be carried out on the *Prolific* platform. The preregistration, R syntax, and clean and raw datasets are accessible at <https://osf.io/yx3tr/>.

Hypotheses

1. There will be a main effect of reduced susceptibility to misinformation for the *Bad News* game
2. There will be a main effect of reduced susceptibility to misinformation for sadness
3. There will be a main effect of increased susceptibility to misinformation for happiness

Method

Participants and Design

The current study is a pre-registered, between-subjects, 2 (treatment, control) x 3 (happy, neutral, sad) double-blinded, randomised controlled factorial design experiment, featuring mood induction procedures (MIPs). The independent variables (between-subjects) were psychological inoculation against misinformation (*Bad News* as treatment and the *Tetris*-style game *Block Puzzle* as control) and mood (happy, sad, neutral). The dependent variable was susceptibility to misinformation as assessed by between-group mean pre-post intervention differences in ratings

of ‘real’ and ‘fake’ news headlines. Social desirability bias was included as a potential covariate to control for possible demand characteristics, and depressive symptoms was included as a screening measure. The Ethics Committee of the Faculty of Behavioural and Social Sciences at the University of Groningen granted ethics approval: PSY-2223-S-0190.

Participants were undergraduate psychology students of the University of Groningen (RUG) who were at least 18 years of age and reported being able to complete the study in English. Students self-enlisted via the SONA online scheduling system, and participated for course credits. The study was conducted in an experimental laboratory of the Faculty of Behavioural and Social Sciences (BSS) at the RUG, between 13 and 21 March 2023. *Qualtrics* facilitated the survey, including random assignment to conditions, the delivery of MIPs and games, and data collection.

A power analysis conducted in G*Power for a fixed effects, one-way ANOVA including six groups revealed a total sample of $N = 216$ would be required to achieve 80% power to detect moderate between-group differences in effects ($f = 0.25$; $\alpha = .05$). A total sample of $N = 250$ was sought, and $N = 248$ achieved. Of the participants, one did not complete the study and two did not report a final score for Bad News. Data pertaining to those participants were excluded, leaving $N = 245$. There was otherwise no missing data and no attrition. The treatment group were 122 participants and the control group 123. Group sizes by mood condition were $n_{\text{happy}} = 82$ ($n_{\text{happy-treat.}} = 41$, $n_{\text{happy-cont.}} = 41$), $n_{\text{neutral}} = 83$ ($n_{\text{neutral-treat.}} = 42$, $n_{\text{neutral-cont.}} = 41$), and $n_{\text{sad}} = 80$ ($n_{\text{sad-treat.}} = 39$, $n_{\text{sad-cont.}} = 41$).

Materials

Susceptibility to Misinformation

We employed the eight-item Misinformation Susceptibility Test (MIST-8; Maertens et al., 2021) to operationalise the dependent variable. The MIST-8 includes the most

psychometrically robust items of the MIST-20, which is a scale comprising of 10 ‘real’ and 10 ‘fake’ news headlines. Participants were asked to rate the headlines as either ‘real’ or ‘fake’ (four items of each). Cronbach’s alpha for the MIST-8 was unacceptable $\alpha = .28$. Cronbach’s alpha can underestimate scale reliability (Revelle & Condon, 2019), so McDonald’s ω (McDonald, 1999) was calculated as an additional indicator of reliability, resulting in pre $\omega = .37$, which was improved by still unacceptable. As such, results will be presented for the four individual ‘fake news’ items of the MIST-8.

Mood

The Joviality and Sadness subscales from the self-report measure Positive and Negative Affect Schedule–Expanded (PANAS-X; Clark & Watson, 1994), assessed mood pre- and post-MIPs. The Joviality subscale includes eight items and the Sadness scale five. The median Participants are asked to ‘Please answer honestly how you feel right now’, on a five-point Likert scale (0 = very slightly or not at all, 4 = extremely) in response to an affective adjective, for example, ‘sad’. Internal consistency was good for the sadness subscale, and excellent for the joviality subscale (Sad $\alpha = .83$; Jovial $\alpha = .93$).

Social Desirability Bias

The Social Desirability Scale (SDS; Stöber, 2001) in its revised form includes 16 items. Participants were directed to answer ‘True’ or ‘False’ to descriptions that are either socially desirable (6) or undesirable (10). A response that frames the respondent in a socially desirable light (e.g., ‘True’ on a desirable description) is coded as 1. Socially undesirable descriptions were reverse coded (1 for ‘False’) so that higher scores indicate higher social desirability bias. The SDS had poor internal consistency ($\alpha = .59$; $\omega = .59$).

Depressive Symptoms

The Patient Health Questionnaire-2 (PHQ-2; Kroenke et al., 2010) is a two-item screening tool assessing depressed mood. As per Marcusson-Clavertz et al., (2019), who validated the mood induction procedures used, any participant that reported little or no pleasure in doing things or feeling down and depressed on ‘more than half of the days’ or ‘nearly every day’ over the past two weeks, failed the screening criteria.

Procedure

The researcher led participants to a cubicle in which a desktop computer was set to the start of a survey hosted in *Qualtrics*. After being informed and giving consent, participants provided demographic information and completed the depressive symptoms screening. They then filled out mood and susceptibility to misinformation pre-tests, and the social desirability scale. Random assignment to conditions followed, along with MIPs and an instruction to play either *Bad News* or *Block Puzzle*. All participants then completed post-test measures of mood and susceptibility to misinformation, then were debriefed on the purpose of the study. Finally, information and support were provided for those who still felt sad, which included an opportunity to watch ‘Hakuna Matata’ from the happy MIP.

Mood Induction Procedures (MIPs)

To induce mood the present study employed MIPs validated for online use by Marcusson-Clavertz et al. (2019). Happy and sad MIPs are identical to those described in the main study, except where noted below.

While participants in the sad condition were playing the intervention or control game they listened to validated mood-inducing music from Fakhrhosseini and Jeon (2017): “At the Ivy Gate” by Brian Crain, “Prelude in E minor, Op. 28”, by Chopin, “Into the Dark”, by Sebastian Larsson, and “Lemminkäinen Suite, Op. 22: No. 2”, by Sibelius.

The neutral MIP featured a video edited from a documentary program on magnets ('Modern Marvels', season 8, episode 35). The clip comprised a vignette in which magnets are introduced, demonstrations of attraction and repulsion and of iron filings on paper, an animation of electrons creating a magnetic field, and an introduction to the compass. The music following was "Variations for Winds, Strings and Keyboards", by Steve Reich, to which participants continued to listen while engaging in gameplay.

While playing the games, happy condition participants listened to the Brandenburg Concerto No. 2 in F Major, and No. 3 in G Major, by J. S. Bach (Fakhrhosseini & Joen, 2017).

Results

Descriptive statistics, assumption tests, manipulation checks and notes regarding the depressive symptoms screening appear after the results for hypothesis testing and a summary.

Hypothesis Testing

Preregistered analyses were not preferred on account of the poor reliability of the scale operationalising the DV and because MIPs effectively delivered two mood groups instead of three (ANOVA no longer required for mood-related hypotheses). The exploratory nature of this pilot study recommended the analyses that would best inform the study to follow should be preferred. As such, robust *t*-tests of independent mean differences-in-difference in T1 to T2 scores on each of the fake items of the MIST-8 were conducted for the conditions of intervention (referent to H1), and mood (referent to H2 and H3). Tables A1-A3 show H1-H3 results.

Hypothesis 1

H1 predicted a main effect of reduced susceptibility to misinformation for the *Bad News* game. Susceptibility to misinformation was operationalised by accuracy of classification of each of the four fake news headlines of the MIST-8. H1 was supported for items 1 and 2, with marginal results for items 3 and 4.

Table A1

Robust Comparisons of Mean T1-T2 Differences in Correct Responses to Fake News MIST Items in the Bad News Condition

Item	$t(121)$	p (one-tailed)	95% CI	M_{DIFF}	d
1	7.84	< .001	(0.26, 0.44)	0.35	0.70
2	3.02	< .01	(0.05, 0.22)	0.13	0.27
3	1.54	.06	(-0.02, 0.13)	0.06	0.14
4	1.68	.05	(-0.01, 0.11)	0.04	0.15

Note. M_{DIFF} = the mean difference between T1 and T2.

Hypothesis 2

H2 predicted a main effect of reduced susceptibility to misinformation for sadness. Susceptibility to misinformation was operationalised by accuracy of classification of each of the four fake news headlines of the MIST-8. H2 was supported for items 1, but not supported for items 2, 3, and 4.

Hypothesis 3

H3 predicted a main effect of increased susceptibility to misinformation for happiness. Susceptibility to misinformation was operationalised by accuracy of classification of each of the four fake news headlines of the MIST-8. H3 was not supported for any items, with significant effects on the opposite direction for items 1 and 2.

Table A2

Robust Comparisons of Mean T1-T2 Differences in Correct Responses to Fake News MIST Items in the Sad Group

Item	$t(79)$	p (one-tailed)	95% CI	M_{DIFF}	d
1	3.53	< .001	(0.08, 0.27)	0.18	0.39
2	0.63	.27	(-0.05, 0.10)	0.03	0.07
3	-0.38	.71 (two-tailed)	(-0.08, 0.05)	-0.01	-0.04
4	0.00	1.00 (two-tailed)	(-0.05, 0.05)	0.00	0.00

Table A3

Robust Comparisons of Mean T1-T2 Differences in Correct Responses to Fake News MIST Items in the Non-Sad (Happy and Neutral) Group

Item	$t(164)$	p (two-tailed)	95% CI	M_{BN}	M_{BP}
1	5.31	< .001	(0.13, 0.27)	0.20	0.41
2	2.30	.02	(0.01, 0.16)	0.08	0.18
3	1.28	.21	(-0.02, 0.09)	0.04	0.10
4	1.27	.21	(-0.01, 0.06)	0.02	0.10

Note: M_{BN} = the mean for the Bad News condition, M_{BP} = the mean for the Block Puzzle condition.

Summary

H1 and H2 were partially supported, while effects were in the opposite direction for H3. Follow up post-hoc difference-in-differences analyses supported a favourable effect for Bad News on the accuracy of reliability rating for fake news headlines, but did not support a differential effect of mood on the same. Results thus provided preliminary support for a positive effect on accuracy in reliability ratings of ‘fake news’ headlines for the inoculation game Bad News, while there was partial support for an effect of sadness in reducing susceptibility to misinformation.

Other outcomes of note were that the MIST-8 was found to lack internal consistency with a considerable ceiling effect for item 4. Further, at least one participant was confused by the framing of ‘fake’ or ‘real’ for news items, unsure whether that referred to veracity or if the item was a genuine example of a news headline. Regarding mood effects, there was no clear difference between the happy and neutral groups, and descriptive differences in post-intervention mood in the gameplaying control condition consistent with a mood repair effect of casual video games (Pine et al., 2020). Finally, the screening for depressive symptoms and measure of social desirability went unutilised as neither were associated with the quality of mood inductions or reports of mood.

Descriptive Statistics

The mean age was 20.21 years ($SD = 2.12$). There were 61 males, 177 females, and 7 non-gender-identified participants. Table A6 presents Pearson’s correlations between independent variables and covariates, and Table A7 presents biserial correlations of dependent variables with independent variables and covariates. Of note is that SDS was uncorrelated with

any variables, and there were no significant correlations between the ‘fake’ MIST-8 items, pre or post intervention, and any other variables.

Tables A8 and A9 show means and standard deviations for measurements of joviality and sadness, respectively, across all conditions, pre and post. In relation to both mood conditions, post-MIP scores adhere to a pattern of highest joviality and lowest sadness for the happy MIP, the opposite for the sad MIP, and neutral mood scores in-between.

Table A6

Correlations Between Independent Variables, Pre and Post, and Potential Covariates

	Sad _{Pre}	Sad _{Post}	Jovial _{Pre}	Jovial _{Post}	SDS	PHQ1	PHQ2
Sad _{Pre}	-						
Sad _{Post}	.59***	-					
Jovial _{Pre}	-.33***	-.15	-				
Jovial _{Post}	-.22**	-.42***	.69***	-			
SDS	-.10	-.10	.10	.09	-		
PHQ1	.37***	.33***	-.26***	-.20***	-.02	-	
PHQ2	.54***	.34***	-.26***	-.20***	-.08	.41***	-

Note. * = $p < .05$ ** = $p < .01$ *** = $p < .001$.

Table A7

Biserial Correlations of Dependent Variables, Pre and Post, with Independent Variables and Potential Covariates

	Sad _{Pre}	Sad _{Post}	Jovial _{Pre}	Jovial _{Post}	SDS	PHQ1	PHQ2
MIST1 _{Pre}	.07	.05	-.10	-.06	-.04	.07	.08
MIST2 _{Pre}	.03	.12	-.12	-.15	.04	.12	.06
MIST3 _{Pre}	.02	.03	.02	0	.06	.03	.05
MIST4 _{Pre}	-.12	-.03	.01	.02	-.03	-.02	-.11
MIST1 _{Post}	.02	.07	-.07	-.09	-.16	.12	.10
MIST2 _{Post}	-.06	-.03	-.07	-.04	-.03	.08	.05
MIST3 _{Post}	.02	-.02	.02	.07	.01	-.04	-.03
MIST4 _{Post}	-.06	-.01	-.13	-.13	-.11	-.01	0

Note. No correlations were statically significant at $\alpha = .05$.

Table A8

Means and Standard Deviations for Joviality Scores Across All Conditions, Pre and Post.

Mood Group	Pre	Post	Condition	Pre	Post M
	M (SD)	M (SD)		M (SD)	M (SD)
Happy	2.93 (0.90)	2.98 (0.93)	Bad News	2.99 (0.86)	2.91 (0.94)
			Block Puzzle	2.88 (0.94)	3.04 (0.93)
Neutral	2.85 (0.87)	2.81 (0.97)	Bad News	2.85 (0.88)	2.79 (0.94)
			Block Puzzle	2.84 (0.87)	2.83 (1.01)
Sad	2.58 (0.87)	2.13 (0.84)	Bad News	2.38 (0.88)	1.92 (0.68)
			Block Puzzle	2.76 (0.83)	2.34 (0.92)

Table A9*Means and Standard Deviations for Sadness Scores Across All Conditions, Pre and Post.*

Mood Group	Pre	Post	Condition	Pre	Post M
	M (SD)	M (SD)		M (SD)	M (SD)
Happy	1.46 (0.49)	1.32 (0.48)	Bad News	1.44 (0.39)	1.35 (0.49)
			Block Puzzle	1.48 (0.58)	1.29 (0.48)
Neutral	1.58 (0.69)	1.43 (0.57)	Bad News	1.66 (0.78)	1.55 (0.66)
			Block Puzzle	1.51 (0.58)	1.30 (0.44)
Sad	1.70 (0.71)	1.83 (0.69)	Bad News	1.70 (0.73)	1.87 (0.69)
			Block Puzzle	1.71 (0.69)	1.79 (0.70)

Means and standard deviations of the raw scores for each fake news item of the MIST-8 are provided in Tables A10-A13. A mean of 0.5 signifies that by aggregate participants were as often incorrect as they were correct in their judgements. Of note is pre-intervention heterogeneity between conditions in ratings of MIST item 1, an apparent ceiling effect for MIST item 4, and a trend for post-intervention groups that played Bad News to score higher than those who played Block Puzzle.

Table A10

Means and Standard Deviations for MIST Item 1 Scores (Fake News), Pre and Post, by Inoculation and Mood Conditions, N = 245

Time	Inoculation Condition	Mean	SD	Mood Condition	Mean	SD
Pre	Bad News	.41	.49	Happy	0.22	0.42
				Neutral	0.55	0.50
				Sad	0.46	0.51
	Block Puzzle	.44	.50	Happy	0.54	0.50
				Neutral	0.41	0.50
				Sad	0.37	0.49
Post	Bad News	.76	.43	Happy	0.73	0.45
				Neutral	0.74	0.45
				Sad	0.82	0.39
	Block Puzzle	.47	.50	Happy	0.59	0.50
				Neutral	0.46	0.51
				Sad	0.37	0.48

Note. higher scores denote lower susceptibility.

Table A11

Means and Standard Deviations for MIST Item 2 Scores (Fake News), Pre and Post, by Inoculation and Mood Conditions, N = 245

Time	Inoculation Condition	Mean	SD	Mood Condition	Mean	SD
Pre	Bad News	.70	.46	Happy	0.68	0.47
				Neutral	0.71	0.46
				Sad	0.72	0.46
	Block Puzzle	.71	.47	Happy	0.68	0.47
				Neutral	0.66	0.48
				Sad	0.78	0.42
Post	Bad News	.84	.37	Happy	0.85	0.36
				Neutral	0.83	0.38
				Sad	0.82	0.39
	Block Puzzle	.71	.46	Happy	0.71	0.46
				Neutral	0.68	0.47
				Sad	0.73	0.45

Note. higher scores denote lower susceptibility.

Table A12

Means and Standard Deviations for MIST Item 3 Scores (Fake News), Pre and Post, by Inoculation and Mood Conditions, N = 245

Time	Inoculation Condition	Mean	SD	Mood Condition	Mean	SD
Pre	Bad News	.83	.38	Happy	0.83	0.38
				Neutral	0.81	0.40
				Sad	0.85	0.37
	Block Puzzle	.76	.43	Happy	0.85	0.36
				Neutral	0.63	0.49
				Sad	0.80	0.40
Post	Bad News	.89	.32	Happy	0.88	0.33
				Neutral	0.86	0.35
				Sad	0.92	0.27
	Block Puzzle	.75	.44	Happy	0.83	0.38
				Neutral	0.71	0.46
				Sad	0.71	0.46

Note. higher scores denote lower susceptibility.

Table A13

Means and Standard Deviations for MIST Item 4 Scores (Fake News), Pre and Post, by Inoculation and Mood Conditions, N = 245

Time	Inoculation Condition	Mean	SD	Mood Condition	Mean	SD
Pre	Bad News	.93	.25	Happy	1.00	0.00
				Neutral	0.86	0.35
				Sad	0.95	0.22
	Block Puzzle	.94	.23	Happy	0.93	0.26
				Neutral	0.95	0.22
				Sad	0.95	0.22
Post	Bad News	.98	.16	Happy	0.98	0.16
				Neutral	0.98	0.15
				Sad	0.97	0.16
	Block Puzzle	.93	.25	Happy	0.90	0.30
				Neutral	0.98	0.16
				Sad	0.93	0.26

Note. higher scores denote lower susceptibility.

Assumption Testing

For MIP checks

The distributions for pre-measures of sadness and jovial for the whole sample were taken as a proxy for the sampling distribution. Sadness presented a floor effect, positive skew ($2/SE = 5.6$) and leptokurtosis ($2/SE = 5.4$). Joviality was approximately normal with mild platy kurtosis ($2/SE = -1.25$). Levene's tests for homogeneity of variances on pre- and post-MIP mood ratings

showed that variances in joviality ratings were not significantly heterogenous between mood induction groups, $F(2, 242) = 0.01$, ns, and $F(2, 242) = 1.85$, ns (for pre and post respectively). Variances in pre-MIP ratings of sadness were not significantly different either: $F(1, 242) = 2.91$, $p = .06$. But for post-MIP ratings of sadness, they were: $F(1, 242) = 5.50$, $p < .01$. The residuals for post intervention joviality ratings were distributed normally, but for ratings of sadness there was a floor effect resulting in positive skew and leptokurtic kurtosis.

For hypothesis testing

The distribution for pre-measures of the MIST-8 fake news items for the whole sample were taken as a proxy for the sampling distribution. The distribution was found to be negatively skewed ($2/SE = -1.88$). Residuals appeared normally distributed. Levene's tests for homogeneity of variances showed variances between treatment and control were homogenous pre-intervention, $F(1, 243) = 1.64$, ns, but not post-intervention, $F(1, 243) = 10.63$, $p < .01$.

The various violations of assumptions for t -tests and ANOVA, assessed within the context of the relatively large sample size and the exploratory nature of the study, recommended that findings should be interpreted with caution.

Screening for Depressive Symptoms Regarding MIP Efficacy

On the two items of the PHQ2 depressive symptoms scale, $n = 49$ scored a 3 or 4 on at least 1 of the items. The pre-registration calls for these participants to be screened out as this was the procedure detailed in (Marcusson-Clavertz et al., 2019). The reason for screening was because a mood induction might not be expected to work on those with depressive symptoms.

Robust repeated measures ANOVAs showed significant differences between the three mood groups on joviality and sadness ($N = 196$). Planned contrasts indicated that the group that did not have a sad mood induction (happy and neutral) reported significantly higher joviality compared to the sad mood induction group, $t(193) = 5.21$, $p < .001$, and joviality was not

significantly different between the happy or neutral mood induction groups, $t(242) = -1.20, p < .12$ (one-tailed). Further, the sad mood induction group reported significantly higher sadness compared to the group that did not have a sad mood induction, $t(242) = -5.66, p < .001$, and sadness was not significantly different between the happy or neutral mood induction groups, $t(242) = -.35, p < .37$ (one-tailed).

Overall, the mood inductions worked at least as well or better in the larger sample, which was retained in the interests of preserving power.

Appendix B

Scales and Adaptations

Table B1

Items of the MIST-20

Please categorise the following news headlines as either 'Reliable News' or 'Unreliable News'.

Some items may look credible or obviously false at first sight, but may actually fall in the opposite category. However, for each news headline, only one category is correct:

#	News Headline
1	Government Officials Have Manipulated Stock Prices to Hide Scandals
2	The Corporate Media Is Controlled by the Military-industrial Complex: The Major Oil Companies Own the Media and Control Their Agenda
3	New Study: Left-Wingers Are More Likely to Lie to Get a Higher Salary
4	The Government Is Manipulating the Public's Perception of Genetic Engineering in Order to Make People More Accepting of Such Techniques
5	Left-Wing Extremism Causes 'More Damage' to World Than Terrorism, Says UN Report
6	Certain Vaccines Are Loaded with Dangerous Chemicals and Toxins
7	New Study: Clear Relationship Between Eye Colour and Intelligence
8	The Government Is Knowingly Spreading Disease Through the Airwaves and Food Supply
9	Ebola Virus 'Caused by US Nuclear Weapons Testing', New Study Says
10	Government Officials Have Illegally Manipulated the Weather to Cause Devastating Storms
11	Attitudes Toward EU Are Largely Positive, Both Within Europe and Outside It
12	One-in-Three Worldwide Lack Confidence in NGOs
13	Reflecting a Demographic Shift, 109 US Counties Have Become Majority Non-white Since 2000
14	International Relations Experts and US Public Agree: America Is Less Respected Globally
15	Hyatt Will Remove Small Bottles from Hotel Bathrooms by 2021
16	Morocco's King Appoints Committee Chief to Fight Poverty and Inequality
17	Republicans Divided in Views of Trump's Conduct, Democrats Are Broadly Critical
18	Democrats More Supportive than Republicans of Federal Spending for Scientific Research
19	Global Warming Age Gap: Younger Americans Most Worried
20	US Support for Legal Marijuana Steady in Past Year

Note. Item responses (categorisations) adapted from 'Fake News' and 'Real News', # = Item number.

Figure B1

The MDMQ

In the following you find a list of expressions that characterize different moods. Please take a look at the list, word by word, and mark for each word the answer that represents best the actual intensity of your mood status.

Example:

Right now I feel ...

definitely not	not really	a little	very much	extremely
1	2	3	4	5

good

Supposed you feel very good at the moment, you would fill out the circle number 5:

Right now I feel ...

definitely not	not really	a little	very much	extremely
1	2	3	4	5

good

Please pay attention to the following facts:

- Within the list there are some attributes that possibly describe the same or similar moods. Please do not get irritated due to this fact, and judge each attribute irrespective of your answer to another attribute.
- Please judge only how you feel at this moment, and not how you normally or sometimes feel.
- If you have some difficulties in finding an answer, please mark those answer that fits best.

Please judge each word and do not leave out a word.

Note. Responses were adapted to ‘Not at all’, ‘Hardly at all’, ‘Not really’, ‘Somewhat’, ‘Quite a bit’, and ‘Very’. The adjectives used at T1 (left side) were ‘Content’, ‘Bad’(reverse coded), ‘Great’, and ‘Uncomfortable’ (reverse coded), and used at T2 (right side) were ‘Good’, ‘Unhappy’ (reverse coded), ‘Discontent’ (reverse coded), and ‘Happy’.

Table B2*The Apprehensive Threat Scale*

Original	Adaptation
<p>The next set of items are designed to help us to understand how you feel about the idea expressed at the beginning of the message you just read that, despite your opinion on this issue, there is a possibility you may come into contact with arguments contrary to your position that are so persuasive that they may cause you to rethink your position. I find this possibility</p>	<p>The next set of items are designed to help us to understand how you feel about the idea that you will come into contact with misleading information online that is so persuasive that it may cause you to take an incorrect position on an important issue. I find this possibility*</p>

Note. Responses on a 7-point Likert scale for 5 binary adjectives pairs:

nonthreatening/threatening, not harmful/harmful, not risky/risky, not dangerous/dangerous, calm/anxious, and not scary/scary.

Table B3*The Motivational Treat Scale*

Indicate your level of agreement with the following statements:

Original Statement	Adapted Statement
I want to defend my current attitudes from attack	I want to defend my current attitudes from attack
I feel motivated to think about why I hold the beliefs I do about 9/11	I feel motivated to think about why I might accept or reject a piece of online information
I feel motivated to resist persuasive messages about alternative accounts of 9/11	I feel motivated to resist persuasive messages communicating misleading accounts of events and issues
I want to counterargue conspiracy theories about 9/11	I want to counterargue techniques for spreading online misinformation

Note. Responses on a 7-point Likert-type scale (1 = strongly disagree; 7 = strongly agree).

Appendix C


Visual Stimuli from the Control Condition Video




1. Impacting choice architecture
Changing the context in which decisions are made

Choice architecture

Creating moments where users reflect before they share, can improve decision-making. For instance, [shifting users' attention to accuracy](#) has been shown to increase the quality of news they share. Additionally, creating friction to encourage pausing and reflecting, social media platforms [have attempted to reduce](#) the spread of mis/disinformation by asking users "Are you sure you want to share this video?" before they share videos with unsubstantiated content.



2. Refutation
Proactively providing warnings, pre-emptive refutations, and/or explaining misleading techniques

Refutation



Warnings and fact-checking

Warnings based on fact checks limit the credibility of untrustworthy messengers and may also decrease the amount of mis/disinformation sharing ([crowdsourcing labels](#) could also be an option). Some social media platforms have attempted to reduce the spread and credibility of mis/disinformation by adding a warning (e.g. "misleading", "truthful", "not verified") to potentially misleading content. While labelling is often employed, issues surrounding [effectiveness](#) and possible [spillover effects](#) remain in question.



Pre-bunking

Once people see and believe mis/disinformation, they tend to stay consistent with their formed beliefs. This tendency stems from a desire to avoid cognitive dissonance, or the uncomfortable state of having two different beliefs that conflict with each other. Cognitive dissonance can be proactively avoided through pre-bunking strategies.

Drawing on the psychological theory known as inoculation theory, pre-bunking takes a proactive approach to deal with mis/disinformation by exposing individuals to small doses of mis/disinformation and by equipping them with the necessary rhetorical and argumentative resources to refute mis/disinformation before they encounter it. Pre-bunking thus provides people with the skills necessary to protect themselves against being misinformed.

Games like [Go Viral](#), [Bad News](#) and [Sweet Victory](#) employ pre-bunking in game form. Having players work to spread misinformation, teaches and reminds them of the tactics that misinformation leverages in the process.



3. Boosting

Introducing education and tools to improve individual decision-making capabilities



Boosting

A key challenge in the current information ecosystem is distinguishing trustworthy sources of information from untrustworthy ones. [Boosts](#) use behavioural science to improve individual decision-making capabilities. There are a range of boosts including educational tools, media literacy tips and fast-and-frugal decision trees. These mental tools can be applied to combat mis/disinformation. More generally, educating users can help shape their default behaviours and promote the formation of positive digital and classic media consumption habits, news interpretation, and sharing.

PLEASE TYPE THE LETTER M IN THE BOX BELOW



Lateral Reading

One boost is the [lateral reading](#) strategy. It teaches people to emulate [how professional fact checkers](#) establish the credibility of online information. Lateral reading involves opening up new browser tabs to seek information about the individual or organisation behind an unfamiliar website before investing significant cognitive effort to process its contents. If the outcome of this source check is negative (i.e., this website is not trustworthy), no further effort needs to be invested. Enhancing or instilling new competencies like lateral reading is essential to ignore information and to analyse information critically.



Education

Educating users on media literacy and critical media consumption is another way to boost individuals' ability to navigate our information ecosystem. For example, there are [reports](#) and [guidelines for teachers and educators](#) on tackling mis/disinformation and promoting digital literacy through education and training. There is also the argument that critical thinking, the cognitive strategy to help identify valid information, needs to be complemented by the competence of [critical ignoring](#) in order to avoid sources of information that should have been ignored.

GLOBAL BEHAVIOURAL SCIENCE EFFORTS

Global collaboration is crucial in mitigating the challenges that mis/disinformation impose on institutional trust, health, safety, and public discourse. As the world's only truly universal global organisation, the United Nations is positioned to help coordinate efforts to address the issue of mis/disinformation and to facilitate the sharing of real-time behavioural science findings between governments, social media companies, civil society and research organisations. Recognising this need, the UN proposed [a draft resolution on disinformation](#) to the General Assembly, which urges social media companies to ensure their business models and processes comply with international human rights standards. The following section provides UN examples of behavioural science applied to mis/disinformation, although work in this area is in its early days.

TACKLING THE ISSUE MORE BROADLY

Although there have been promising results for some of the outlined approaches, challenges have arisen which [merit further exploration](#). These include the scalability, [replicability](#) and level of impact of behavioural science interventions, including the magnitude and degree of sustained effect. Furthermore, while behavioural science can help address these issues, it should be seen as just one tool in a broader toolbox of interventions that can be used in combination with other solutions. For instance, behavioural science could be leveraged as a flexible, less restrictive, and potentially more proactive complement to regulation and corrective solutions.

PLEASE TYPE THE NUMBER 7 IN THE BOX BELOW

This brief was prepared by the UN Behavioural Science Group of the UN Innovation Network with the support of the Executive Office of the Secretary-General and in collaboration with UN Entities, Member States, academics among other partners. The Group brings together UN colleagues interested in applying behavioural science and welcomes non-UN colleagues in an observer role. [Join us!](#)

www.UNBehaviouralScience

[@behavioural-science@uninnovation.network](mailto:behavioural-science@uninnovation.network)

twitter.com/UN_BeSci

Appendix D

Main Study Descriptive Statistics, Assumption Testing, and Mood Manipulation Checks

Descriptive Statistics

The mean age was 41.71 years ($SD = 12.58$; range = 18-68 years). Of the 368 participants that were invited to contribute at phase two there were 163 males, 204 females, and 1 non-gender-identified person. There were no missing data at either phase.

Table D1 presents Pearson's correlations between all variables. Of note was that post the MIPs at T2, mood had a weak-to-moderate negative relationship with apprehensive threat, but was uncorrelated with other variables. Apprehensive threat had a moderate-to-strong positive association with motivational threat, and a weak-to-moderate negative association with scores on the MIST-20 at T3, but was otherwise unrelated to variables of interest. Motivational threat, on the other hand, had a positive weak-to-moderate relationship with scores on the MIST-20 at all timepoints, and with $MIST_{REAL}$ at T1, but no relationship with $MIST_{FAKE}$ at any timepoint. The pattern of correlations between $MIST_{REAL}$ and $MIST_{FAKE}$ showed they were predominantly unrelated, and only weakly where they were, despite both being strongly correlated with the full MIST-20 of which they are a part. Finally, there was a positive relationship between age and scores on the MIST-20 and $MIST_{REAL}$ at T2 and T3.

Table D2 shows means and standard deviations for mood scores across all conditions, pre and post inductions. The mean rating out of 35 for perceived apprehensive threat in the full sample was 24.20 ($SD = 7.6$). In the happy group it was 24.15 ($SD = 7.00$), and 24.65 ($SD = 7.31$) in the sad. Motivational threat was rated by the full sample at 22.22 ($SD = 3.13$), out of a possible 28. 22.24 ($SD = 3.12$) was the happy group's rating and 22.20 ($SD = 3.14$) the sad.

Table D1*Correlations Between All Variables, T1-T3.*

	Age	Mood T1	Mood T2	App.	Mot.	MIST T1	Fake T1	Real T1	MIST T2	Fake T2	Real T2	MIST T3	Fake T3
Age													
Mood T1	-.01												
Mood T2	.01	.64***											
App.	.00	-.03	-.18**										
Mot.	.01	.12	.04	.41***									
MIST T1	.14	-.03	-.03	.04	.24***								
Fake T1	.11	.04	.00	.03	.11	.72***							
Real T1	.11	-.08	-.03	.03	.24***	.81***	.18**						
MIST T2	.16*	-.01	.01	.00	.16*	.73***	.51***	.61***					
Fake T2	.02	-.03	-.04	.13	.09	.54***	.73***	.15*	.54***				
Real T2	.17*	.01	.03	-.09	.13	.48***	.09	.61***	.81***	-.06			
MIST T3	.22***	-.01	.03	-.18**	.18**	.77***	.53***	.65***	.81***	.43***	.66***		
Fake T3	.08	-.01	-.01	.14	.14	.54***	.76***	.15*	.45***	.80***	-.01	.53***	
Real T3	.21**	-.01	-.04	.12	.12	.57***	.16*	.68***	.68***	.01	.79***	.85***	.00

Note. * = $p < .05$, ** = $p < .01$, *** = $p < .001$. App. = Apprehensive Threat, Mot. = Motivational Threat, MIST = MIST-20, Fake = MIST_{FAKE}, Real = MIST_{REAL}

Table D2*Means and Standard Deviations for Mood Scores Across All Conditions, Pre and Post.*

Mood Group	T1	T2	Condition	T1	T2
	M (SD)	M (SD)		M (SD)	M (SD)
Happy	17.78 (3.62)	18.08 (3.75)	Go Viral!	17.65 (3.87)	17.58 (4.06)
			Control	17.90 (3.38)	18.55 (3.37)
Sad	17.45 (3.79)	15.38 (4.36)	Go Viral!	17.68 (3.93)	16.10 (4.40)
			Control	17.23 (3.65)	14.66 (4.22)

Tables D3 and D4 show apprehensive and motivational threat, respectively, across the remaining conditions.

Table D3*Means and Standard Deviations for Apprehensive Threat*

Inoculation Condition	Mean	SD	Mood Group	Mean	SD
Go Viral!	25.60	6.75	Happy	24.85	7.13
			Sad	26.32	6.32
Control	23.23	7.36	Happy	23.48	6.86
			Sad	22.97	7.86

Table D4*Means and Standard Deviations for Motivational Threat*

Inoculation Condition	Mean	SD	Mood Group	Mean	SD
Go Viral!	22.38	3.25	Happy	22.51	3.10
			Sad	22.27	3.41
Control	22.06	3.00	Happy	21.99	3.12
			Sad	22.14	2.87

Means and standard deviations for MIST-20 scores by inoculation condition and mood over all time points are provided in Table D5. Means and standard deviations for MIST_{FAKE} ($n = 10$) by inoculation condition and mood over all time points are provided in Table D6, and for MIST_{REAL} ($n = 10$) in Table D7.

Table D5

Means and Standard Deviations for MIST-20 Scores by Condition, Across Timepoints.

Timepoint	Inoculation Condition	Mean	SD	Mood Group	Mean	SD
T1	Go Viral!	15.42	3.24	Happy	15.91	2.89
				Sad	14.95	3.48
	Control	15.83	2.77	Happy	16.34	2.72
				Sad	15.32	2.75
T2	Go Viral!	15.01	3.35	Happy	15.52	3.39
				Sad	14.52	3.25
	Control	15.38	3.20	Happy	15.82	3.32
				Sad	14.94	3.04
T3	Go Viral!	15.31	3.34	Happy	15.76	3.17
				Sad	14.85	3.48
	Control	15.89	3.32	Happy	16.00	3.37
				Sad	15.17	3.23

Note. A mean of 10 signifies that, by aggregate, participants were as often incorrect as they were correct in their judgements of real and fake items. Higher scores denote better discernment.

Table D6

Means and Standard Deviations for MIST_{FAKE} Scores by Condition, Across Timepoints.

Timepoint	Inoculation Condition	Mean	SD	Mood Condition	Mean	SD
T1	Go Viral!	8.03	1.92	Happy	8.26	1.72
				Sad	7.81	2.08
	Control	8.47	1.67	Happy	8.73	1.45
				Sad	8.22	1.83
T2	Go Viral!	8.53	1.98	Happy	8.58	2.05
				Sad	8.48	1.93
	Control	8.70	1.70	Happy	8.81	1.64
				Sad	8.60	1.76
T3	Go Viral!	8.64	1.77	Happy	8.76	1.67
				Sad	8.53	1.88
	Control	8.72	1.77	Happy	8.86	1.69
				Sad	8.57	1.84

Note. A mean of 5 signifies that, by aggregate, participants were as often incorrect as they were correct in their judgements of fake news items. Higher scores denote better discernment.

Table D7

Means and Standard Deviations for MIST_{REAL} Scores by Condition, Across Timepoints.

Timepoint	Inoculation Condition	Mean	SD	Mood Condition	Mean	SD
T1	Go Viral!	7.39	2.28	Happy	7.65	2.17
				Sad	7.14	2.36
	Control	7.36	2.19	Happy	7.61	2.13
				Sad	7.11	2.23
T2	Go Viral!	6.47	2.82	Happy	6.93	2.73
				Sad	6.03	2.85
	Control	6.67	2.93	Happy	7.01	2.85
				Sad	6.33	2.97
T3	Go Viral!	6.67	2.83	Happy	7.00	2.63
				Sad	6.33	3.00
	Control	6.87	2.82	Happy	7.14	2.74
				Sad	6.60	2.89

Note. A mean of 5 signifies that, by aggregate, participants were as often incorrect as they were correct in their judgements of real news items. Higher scores denote better discernment.

Assumption Testing

The distributions of variables were taken as a proxy for the sampling distribution to assess normality. Results are presented in Table D8. The distributions of age and mood at T1 had mild negative kurtosis. Apprehensive threat and both MIST subscales suffered negative skew and positive kurtosis, while the distributions for motivational threat and the full MIST-20 were negatively skewed.

Table D8*Skew and Kurtosis for Distributions of Age, Threat, and Variables Measured at T1*

Variable	Skew (2/SE)	Kurtosis (2/SE)
Age	< 1	-1.72
Mood T1	< 1	-1.72
Apprehensive Threat T2	-3.21	5.92
Motivational Threat T2	-2.04	< 1
MIST-20 T1	-3.26	< 1
MIST _{FAKE} T1	-5.23	3.91
MIST _{REAL} T1	-3.86	1.38

Visual checks of quantile-quantile plots provided a test for the normal distribution of residuals for all variables at all applicable timepoints. Plots for the MIST_{FAKE} at all timepoints and the MIST_{REAL} at T1 described a positive curve, denoting non-normal residuals for these distributions consistent with the negative skew detailed in Table D8.

Levene's test assessed the homogeneity of variances between conditions and outcome measures for susceptibility to misinformation and perceived threat. Variances for the combined mood and intervention conditions on the MIST-20 at T1 were significantly different, $F(3, 364) = 3.35, p < .05$, as were variances on MIST_{FAKE} at T1 between the mood conditions, $F(1, 366) = 6.34, p < .05$. Other variances were not significantly different, although those between intervention conditions on the MIST-20 at T1 were marginal, $F(1, 366) = 4.02, p = .05$.

To address the violations of assumptions robust measures were favoured where possible, and an alpha of .01 adopted for significance tests associated with signal detection analyses

(Wilcox, 2005). The relatively large sample size meant that data transformations were not considered necessary for t -tests and ANOVAs, thus avoiding issues of interpretation.

Mood Manipulation Checks

Robust t -tests of dependent means presented in Table D9 show that there was no change in mood in the happy group that played Go Viral!, although mood did shift significantly in the expected directions for the other conditions.

Table D9

Pre- versus Post-Mood t -tests of Dependent Means Across Conditions, Bootstrapped 2000 Times and Based on 20% Trimming.

Condition	Y_t	95% CI	p	r
Happy/Go Viral!	-0.24	(-1.12, 0.65)	.61 (two tailed)	.02
Happy/Control	0.53	(-0.03, 1.09)	< .05 (one-tailed)	.27
Sad/Go Viral!	-2.05	(-2.98, -1.13)	< .001 (one-tailed)	.42
Sad/Control	-3.04	(-3.77, -2.30)	< .001 (one-tailed)	.60

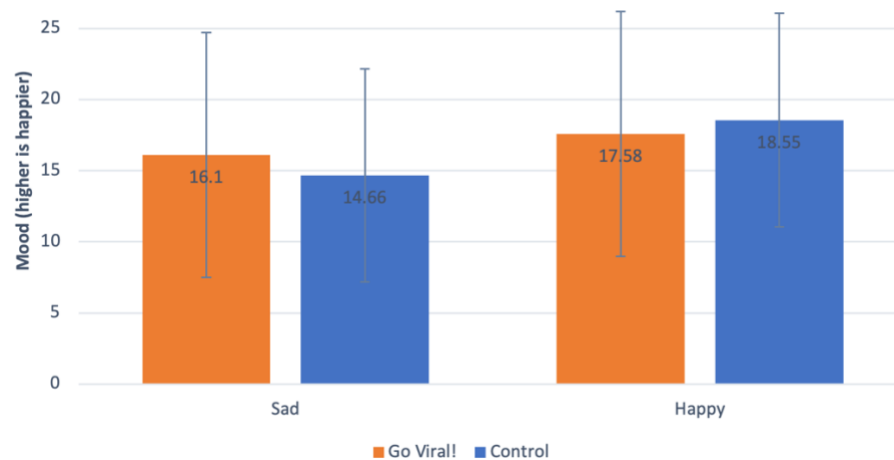
Note. Y_t = the trimmed, bootstrapped difference in means.

A one-way ANOVA not assuming equal variances indicated group differences in mood at T2, $F(3, 200.87) = 17.93, p < .001$. Planned contrasts showed that the intervention and control groups did not differ significantly on mood, $t(364) = 0.57, p = .29$ (one-tailed), and that the happy Go Viral! group was in a significantly better mood at T2 than the sad Go Viral! group, $t(364) = 2.49, p < .01$ (one-tailed). Two additional robust t -tests of independent means comparing mood between intervention conditions within mood groups, with a Bonferroni correction applied for an alpha of .025 (.05/2), revealed no significant differences: for control

verses Go Viral! conditions in the happy group, $\tau(103.17) = 1.10$, $p = .27$, and for control verses Go Viral! conditions in the sad group, $\tau(105.75) = 2.12$, $p = .04$. The overall pattern of results suggests that MIPs were successful in that there was a clear separation between happy and sad groups, and no significant differences within mood groups by intervention condition (although for conditions in the sad group the difference was significant before accounting for the inflated familywise error rate). However, mood inductions were apparently not as effective for those in the happy group, nor for those who played the game. Figure D1 shows T2 mood ratings by condition.

Figure D1

Mood Ratings at T2 by Condition



Note. Error bars show approximate 95% confident intervals.

The relatively lower effect of the happy MIP compared to the sad was anticipated in that it reflected the results of the pilot study which informed the decision to forgo the inclusion of a neutral mood group. But the possible influence of the inoculation intervention implied by descriptive differences in mood by intervention, was unexpected. As participants who played Go

Viral! may have taken longer than ten minutes to play the game and thus exhausted the mood inducing music provided, and as any participant may have abandoned MIP stimuli for a time during the study, time taken to complete the survey was considered as a possible factor affecting mood at T2. A Welch's two sample *t*-test indicated that participants in the Go Viral! condition took significantly longer ($M = 30:52$, $SD = 9:34$) than those who watched the 5-minute control video ($M = 25:05$, $SD = 10:18$), $t(360.47) = 5.77$, $p < .001$. All participants who spent over 15 minutes playing the game (expected time was five minutes), or over 10 minutes on the control video or the MIP video or music track (MIP stimuli were four minutes), were removed ($n = 46$). Two additional participants were also removed: one who played Go Viral! for 14:28 and returned incompatible responses on happiness and unhappiness at T2, and another that skipped the MIP video as soon as possible and lodged a similarly implausible mood rating at T2. The reduced sample was $N = 320$ ($n_{\text{HAPPY-TREAT.}} = 67$, $n_{\text{HAPPY-CONT.}} = 75$, and $n_{\text{SAD-TREAT.}} = 90$, $n_{\text{SAD-CONT.}} = 88$; completion time: $M_{\text{TREAT.}} = 27:51$, $SD_{\text{TREAT.}} = 5:49$; $M_{\text{CONT.}} = 24:20$, $SD_{\text{CONT.}} = 9:26$). MIP checks with the reduced sample did not result in improvements in mood inductions, and did not change the pattern of results except to render the T1-T2 effect of the MIP within the happy control group statistically insignificant. With reducing the sample in this way providing no apparent benefits to the quality of the MIPs, the full sample was retained in the interests of preserving statistical power. MIP results pertaining to the reduced sample are presented in Appendix E.

Appendix E

Mood Manipulation Checks with Reduced Sample ($N = 320$)

Table E1 shows means and standard deviations for mood scores across all conditions, pre and post inductions. Robust t -tests of dependent means presented in Table E2 show that there was no change in mood in the happy group that played Go Viral!, but mood shifted in the expected directions for the other conditions.

Table E1

Means and Standard Deviations for Mood Scores Across All Conditions, Pre and Post.

Mood Group	T1	T2	Condition	T1	T2
	M (SD)	M (SD)		M (SD)	M (SD)
Happy	17.93 (3.48)	18.26 (3.66)	Go Viral!	17.81 (3.72)	17.76 (4.05)
			Control	18.02 (3.32)	18.63 (3.32)
Sad	17.35 (3.83)	15.30 (4.32)	Go Viral!	17.60 (3.98)	16.04 (4.50)
			Control	17.13 (3.71)	14.67 (4.08)

Table E2

Pre- versus Post-Mood t -tests of Dependent Means Across Conditions, Bootstrapped 2000 Times and Based on 20% Trimming, $N = 320$.

Condition	Y_t	95% CI	p	r
Happy/Go Viral!	-0.03	(-1.16, 1.01)	.89 (two-tailed)	.01
Happy/Control	0.48	(-0.12, 1.08)	.06 (one-tailed)	.25
Sad/Go Viral!	-2.00	(-3.03, -0.97)	< .001 (one-tailed)	.41
Sad/Control	-2.93	(-3.66, -2.19)	< .001 (one-tailed)	.60

A one-way ANOVA not assuming equal variances indicated group differences in mood at T2, $F(3, 167.58) = 18.62, p < .001$. Planned contrasts showed that the intervention and control groups did not differ significantly on mood, $t(316) = 0.56, p = .29$ (one-tailed), and that the happy Go Viral! group was in a significantly better mood at T2 than the sad Go Viral! group, $t(316) = 2.57, p < .01$ (one-tailed). Two additional t -tests of independent, 20% trimmed means comparing mood between intervention conditions within mood groups, with a Bonferroni correction applied and a resulting alpha of .025 (.05/2), revealed no significant differences: for control verses Go Viral! conditions in the happy group, $\tau(75.47) = 0.83, p = .41$, and for control verses Go Viral! conditions in the sad group, $\tau(80.99) = 1.97, p = .05$. As with the full sample, the overall pattern of results suggests that mood inductions were not as effective for those in the happy group, nor for those who played the game. However, MIPs were slightly less successful in the reduced sample in that the pre-post difference in mood in the happy control group was statistically insignificant.

There are no consequential benefits for mood inductions by removing participants who took an especially long time in engaging with MIP or intervention media.