Correcting Statistical Misinformation About Scientific Findings in the Media: Causation Versus Correlation

Dulcie Irving[1], Robbie W. A. Clark[1], Stephan Lewandowsky[1,2], Peter J. Allen[1]

[1] School of Psychological Science, University of Bristol, UK

[2] School of Psychological Science, University of Western Australia, Perth, Australia

Author Note

Dulcie Irving https://orcid.org/0000-0002-6005-8648

Robbie W. A. Clark https://orcid.org/0000-0002-2160-313X

Stephan Lewandowsky https://orcid.org/0000-0003-1655-2013

Peter J. Allen https://orcid.org/0000-0002-9690-1545

The pre-registered study protocol, materials and anonymized data are publicly available on the Open Science Framework at https://osf.io/ps98w/. We have no funding or conflicts of interest to disclose.

Correspondence concerning this article should be addressed to Dulcie Irving. School of Psychological Science, University of Bristol, 12a Priory Road, Bristol, BS8 1TU, United Kingdom. Email: ob18285@alumni.bristol.ac.uk

Abstract

Although retractions significantly reduce the number of references people make to misinformation, retracted information nevertheless persists in memory, continuing to influence reasoning. One hundred and twenty-nine lay participants completed an adaptation on the traditional continued influence paradigm, which set out to identify whether it is possible to debunk a piece of common statistical misinformation: inappropriate causal inference based on a correlation. We hypothesised that participants in the correction condition would make fewer causal inferences (misinformation) and more correlational inferences (correction) than those in the no-correction condition. Additional secondary hypotheses were that the number of references made to the misinformation and correction would be moderated by the level of trust in science and scientists, and the amount of television that participants watch. Although the secondary hypotheses were not supported, the data strongly supported the primary hypotheses. This study provides evidence for the efficacy of corrections about misinformation where correlational evidence has been inappropriately reported as causal.

Keywords: Misinformation, Causation/Correlation, Debunking, Motivated Reasoning, Media

Public Significance Statement

A high exposure to statistical misinformation coupled with low levels of statistical literacy leaves the public vulnerable to misleading claims about scientific findings in the media. This study showed that it is possible to successfully debunk misinformation where correlational evidence has been inappropriately reported as causal using corrective messaging. These findings encourage the use of corrective messaging in the media to limit the spread of scientific misinformation and its consequences.

Correcting statistical misinformation about scientific findings in the media: Causation versus correlation

In today's media landscape, a wealth of information is immediately accessible. However, much of this information is inaccurate. Misinformation is a term used to describe information that is initially believed to be valid but later shown to be false, within which falls "disinformation", referring to false information that is deliberately disseminated with the intention to misinform (Lewandowsky et al., 2013). On a micro level, the spreading of misinformation has been shown to have consequences for individual health decisions on immunisation (Pluviano et al., 2017). This in itself has become a significant concern over the last year, as belief in misinformation has been linked with vaccine hesitancy, as well as decreased compliance with government guidelines which are intended to slow the spread of COVID-19, such as mask wearing (Loomba et al., 2021; Roozenbeek et al., 2020; Sallam et al., 2020). On a macro level, misinformation has been implicated in influencing societal attitudes towards climate change, as well as prompting increased political polarization (Berinsky, 2017; Nyhan & Reifler, 2015; van der Linden et al., 2017). These issues clearly demonstrate the problem that dissemination of misinformation poses globally, undermining public trust in science and official authorities, and the need to combat it through means such as debunking.

Media outlets are often guilty of misusing science to attract readers, misinforming and spreading fear in the process (Battley, 2019). One common example of this is popular media articles misrepresenting correlational scientific findings as causal (Adams et al., 2017; Bleske-Rechek et al., 2015). The general lack of statistical understanding and skills of non-expert audiences to critically evaluate scientific claims made in the media leaves the public vulnerable to being misinformed about important issues that rely on the accurate communication of scientific research (Sinatra et al., 2014). Even among researchers, inferential interpretations of correlational data can occasionally extend beyond what is warranted by the evidence (Ksir & Hart, 2016; Sumner et al., 2014). For instance, Sumner and colleagues (2014) found that in their examination of 462 academic press releases, often used to communicate scientific research to the media, 33% contained exaggerated causal claims made from correlational research. Despite laypeople's high exposure and vulnerability to misleading scientific claims in popular media (Scheufele & Krause, 2019), no research is known to us that has investigated the effects of correcting statistical misinformation in the lay population.

Public consumption of statistical information communicated by governments and other official bodies has increased considerably amid the current global coronavirus pandemic as a greater emphasis is placed on science communication (Matta, 2020). A recent study showed that the way in which statistical information in vaccine-related statements was framed to pro-vaccination individuals influenced perceived plausibility and willingness to communicate them to others (Altay & Mercier, 2020). Given the volume of statistical information currently being presented to and consumed by the public, and the impact it can have on attitudes and decisions, there is substantial practical significance in examining whether statistical misinformation, such as misplaced causal interpretations of correlations, can be corrected.

The field of research on correcting misinformation investigates the efficacy of corrections in reducing a reliance on the misinformation, also known as debunking (Chan et al., 2017; Lewandowsky et al., 2017). The most common experimental paradigm used to understand misinformation and its correction is known as the 'warehouse fire', or continued influence paradigm (Johnson & Seifert, 1994; Wilkes & Leatherbarrow, 1988). Its core feature is that participants read a sequence of individual messages which comprise a news story about an event. After reading the story there is a small distractor task, followed by a series of inference and fact-recall questions that measure which information has been remembered. In the case of the warehouse fire, the control group read a description of events without any explanation of how the fire started. The misinformation group read the same story, plus a message providing an incorrect explanation of the fire's origin. The correction group read this same story with the misinformation message, followed later by a correction message, stating that the previously stated cause was incorrect, and that the actual cause has been discovered. In this paradigm, the misinformation effect refers to the extent to which a misinformation message alters belief about a topic, relative to the baseline population level (misinformation vs. control group). The debunking effect reflects the efficacy of a correction message: how it reduces references to the misinformation (correction vs. misinformation group). The misinformation-persistence effect measures the inefficacy of a correction. That is, the number of references that are made to the misinformation in spite of the correction, relative to the baseline population level (correction vs. control group).

Given the novel context, the interest of the current study lay in whether statistical misinformation can be successfully corrected; namely, the debunking effect. A meta-analysis conducted by Chan et al. (2017) showed that debunking effect sizes were large, with a meta-analytic average of *ds* = 1.14-1.33. However, the range of reported effect sizes was much greater, suggesting a certain level of heterogeneity relating to studies of this nature (Walter & Murphy, 2018; Walter & Tukachinsky, 2019).

In the present study we applied the continued influence paradigm, which has traditionally been used to examine general misinformation, to a novel context. We investigated whether it is possible to correct a common form of statistical misinformation present in popular media: inappropriately drawing causal conclusions from correlational evidence. Participants were randomised to one of two experimental conditions: no-correction or correction. They read a fictional news story about the relationship between extended TV watching and cognitive decline, inspired by an article in *The New York Times* (Bakalar, 2019). Informed by previous research, we designed the correction to be as powerful as possible. We therefore included an alternative explanation, in recognition of the fact that individuals prefer to maintain a complete but incorrect model of an event until they are given an alternative explanation to sufficiently fill the gap left by a simple negation (Lewandowsky et al., 2012). Similarly, we ensured that the correction was from a credible source, that it maintained coherency with the story, and explained why the misinformation was inaccurate (Lewandowsky et al., 2012). The primary, confirmatory hypotheses were that participants in the correction condition would make fewer causal inferences (i.e., rely on the misinformation) and more correlational inferences (i.e., rely on the correction) than those in the no-correction condition, in response to the coded inference questions.

As we aimed this study at a non-expert population in a statistical context, we were interested in whether the debunking effect would be influenced by a lack of statistical knowledge. Therefore, in addition to the inference, fact-recall and manipulation-check questions traditionally used in this paradigm, we included a novel statistical knowledge question at the end to ascertain whether participants knew the difference between causal and correlational evidence. This was for exploratory purposes.

Secondary, exploratory hypotheses investigated potential moderating influences of individual differences relating to worldview consistency, grounded in theories of motivated reasoning. This account argues that corrections are less effective, and that misinformation is strengthened, when the correction is contradictory to participants’ pre-existing beliefs (Ecker

& Ang, 2019; Lewandowsky et al., 2012). Worldview backfire effects have been observed in several studies, but are not universal (Hart & Nisbet, 2012; Lewandowsky et al., 2013; Wood & Porter, 2019). We hypothesised that the debunking effect would be moderated by participants' level of trust in science and scientists. Participants with higher levels of trust in science and scientists would show a larger debunking effect than those with lower trust in science and scientists. Trust in science has been discussed in a number of works investigating the efficacy of corrections and worldview consistency (Cook & Lewandowsky, 2016; Lewandowsky et al., 2017; Hyman & Jalbert, 2017). However, it has not yet been measured in this context despite issues with general public trust in science having been highlighted in other research (Achterberg et al., 2017). Given that statistics are a primary mode of scientific communication, and that worldview backfire effects have occurred when scientific subject matter was used, trust in science has particular relevance here (Cook et al., 2017; Lewandowsky et al., 2017).

Secondly, we hypothesised that the amount of time participants spent watching TV might moderate their receptivity to the misinformation and/or correction. Previous research has found that strong beliefs and levels of emotivity can influence misinformation effects and persistence (Ecker et al., 2014; Flynn et al., 2017). The thought that one's TV watching habits could be causing cognitive decline may well be an emotive topic. Therefore, we hypothesised that a) participants who watch more TV would make fewer causal inferences than others in their respective condition (thus avoiding an emotionally challenging conclusion), and b) participants in the correction condition who watch more TV would make more correlational inferences than others in their group.

Method

Participants

Participants were recruited through convenience and snowballing techniques using online methods. The study was distributed to email, Facebook and WhatsApp contacts of the researchers and social media posts inviting people to take part and share the study were made on Facebook. No participation incentives were given and people were assured that non-participation would not affect their relationship with the researchers or the University. Participants completed the study in an online setting and information regarding the location of participants was not collected. However, based on our sampling methods, it is reasonable to assume that the majority were UK residents at the time. One hundred and fifty-seven people started the online experiment, but 28 were removed due to missing data and a failure to reach the debriefing page. Five participants reached the debrief but did not click the submit button. As they had proceeded past all answerable parts of the study this was assumed to be a simple oversight, and so they were included in the analyses.

A final sample of 129 were used to measure the debunking effect (47 males, 82 females, M_{age} = 44.76, *SD* = 16.7). They represented a very well-educated population, with 49.6% having a bachelor's degree, 24.8% a master's degree, 4.7% a doctorate. Just 20.9% had achieved A-levels or less (equivalent to AP examinations in the US). There were two non-native English speakers, both of whom were included as they reported the highest proficiency. All participants completed an informed consent form before taking part.

An *a priori* power analysis was conducted using G*Power (v.3.1; Faul et al., 2009) to determine the required sample size to measure the debunking effect. It suggested a sample size of 150 would be needed to detect an effect as small as d = .54 in a one-tailed independent samples *t*-test, given α = .05 and 1 - β = .95. The estimate effect size d = .54 was chosen because it was the smallest debunking effect size out of the six studies closely resembling our design, which were reported in a meta-analytic review of misinformation correction studies (Chan et al., 2017). We decided upon the lowest value, rather than the median, due to concerns about inflated effect size estimates in non-preregistered published psychology research (Szucs & Ioannidis, 2017). Given that our actual sample size was 129, our actual power for detecting d = .54 or larger was 1 - β = .92.

Design

The primary hypotheses were tested by randomly assigning participants to one of two conditions: correction and no-correction. A third condition in which participants received neither the misinformation nor correction, traditionally used to measure misinformation-persistence, was not included. It would have made little sense to include this baseline condition in this context as there is no obvious third alternative to causal or correlational evidence. 'No' evidence would beg the question, why the story? The two dependent variables were the sum of causal inferences (misinformation) and sum of correlational inferences (correction). The exploratory hypotheses required two additional continuous measured independent variables: trust in science and scientists and average TV watching time. The study was approved by The School of Psychological Science Research Ethics Committee at the University of Bristol (reference number: 86623). The preregistered study protocol, materials and anonymized data are publicly available on the Open Science Framework at: https://osf.io/ps98w/.

Materials

Primary Hypotheses

Materials for the primary hypotheses were largely based on the paradigm used by a small number of related prior studies (Ecker & Ang, 2019; Ecker et al., 2010, 2015; Ecker, Lewandowsky, & Apai, 2011; Ecker, Lewandowsky, Swire, et al., 2011; Johnson & Seifert, 1994; Wilkes & Leatherbarrow, 1988). The fictitious news story consisted of 10 short messages (excluding headline) each presented on a separate page. The story was about the relationship between extended TV watching and cognitive decline. Each message was presented on a separate page. Participants read one of two versions of the story depending on the condition to which they were randomly assigned. For both conditions, Message 4 contained the misinformation: '*watching television for more than 3.5 hours a day causes an increased rate in cognitive decline'*. For the no-correction condition, Message 9 simply continued the story (*The lead author of the study could not be reached for comment*). In the correction condition, Message 9 contained a correction of the initial misinformation which explained that the evidence was correlational rather than causal, also providing an alternative explanation (*The lead author has since specified that the results have been misrepresented and that TV screen time has not yet been found to cause cognitive decline. Only a correlation has been found between the two and no causal link is stated in the study. It is equally plausible cognitive decline is responsible for an extended time spent watching television, or perhaps a third, unknown variable plays a role, such as age*). The remaining messages were fillers that contributed to the overall narrative (Ecker, Lewandowsky, Swire, et al., 2011).

In line with previous uses of this paradigm, participants' comprehension of the story and awareness of the correction were assessed using a post-story questionnaire. Inference questions aimed to assess causal or correlational inferences made by participants in relation to the misinformation and correction, of which five were coded:

1. Based on the news story, why is binge watching TV seen as bad?
2. From what you have just read, what do you think about the relationship between TV and cognitive decline?

3. Based on the results of this study, can we conclude that watching TV causes cognitive decline?
4. From the news story, what effect, if any, do you think extended TV time has on cognition?
5. According to the news story, were TV providers right to be criticised for promoting binge watching? If so, why?

The coded questions explicitly asked participants to answer based on what they had just read, as we wanted to measure inferences rather than beliefs. Given the novel context, we employed a small focus group to improve questions by limiting room for ambiguous or off-topic responses and prepare a coding scheme by pre-empting the range of possible answers.

Ten multiple-choice fact-recall questions followed to determine participant comprehension and served as exclusion criteria (e.g., *How old were the participants used in the study?*). The single open-ended manipulation-check was used to gauge participants' awareness of the correction in the correction condition (*Was any of the information in the story subsequently altered or corrected? If so, what was it?*; Ecker & Ang, 2019). The statistical-knowledge item was '*In your own words, what is the difference between a causal and correlational link?'*. Questions were each presented on a separate page and in the same order for all participants.

Secondary Hypotheses

There were three questionnaires placed before the story that related to the secondary hypotheses. Average TV watching time was measured using one free-text coded question (*On average, how many hours a day to you watch TV?)* embedded within a free-time habits questionnaire that was developed for this study. Trust in science and scientists was measured using a 21-item validated inventory developed by Nadelson et al. (2014), which showed high internal consistency (α = .88). It included items such as '*Scientific theories are trustworthy'.* Participants rated their agreement with these statements on five-point scales ranging from 1 = *'strongly disagree'* to 5 = *'strongly agree'*. Twelve items were reverse coded prior to analysis (Nadelson et al., 2014). An additional spirituality measure (Delaney, 2005) was included to avoid sensitising participants to the true purpose of the study, but responses to this were not coded or included in any analyses.

Procedure

Participants followed a link to an online Qualtrics survey. After reading the information sheet and providing informed consent, some demographic information was collected. Participants then completed the pre-story questionnaire, consisting of three short questionnaires measuring TV watching time, trust in science and scientists, and spirituality. Following this, participants proceeded to the main part of the study, in which they were randomised to the correction or no-correction condition and read the aforementioned messages at their own pace. Participants were informed that they would be asked to recall information about them afterwards and would not be able to re-read the messages. Afterwards, participants completed the post-story questionnaire, which included the inference questions, fact-recall questions, manipulation check, and statistical knowledge check.

Coding

The goal of coding was to ascertain whether participants made a causal (misinformation) or correlational (correction) inference to the five inference questions. Given the novel context, the pre-registered coding scheme did not cover all possible responses. Any additional coding decisions were made using the established principle in the literature of only coding items in the affirmative if they were ‘unequivocally identified’ as being a reference to either source of information (Wilkes & Leatherbarrow, 1988). Responses were coded collaboratively by the first two authors, who were blind to conditions when coding questionnaires, and any disagreements were resolved through discussion. As the data were not coded completely independently, an inter-rater reliability score was not calculated.

Five of ten inference questions were coded twice for participants, once for each of the dependent variables. A participant who made a correlational inference in response to a question would receive a score of 0 under causal inference and 1 under correlational inference, and vice versa. Off topic answers were given a score of 0 on both dependent variables. Confirmatory analyses were conducted on a ‘strict’ coding scheme where arguably ambiguous inferences were scored as 0 on both dependent variables, as were responses that contained ambiguous language (i.e., *linked, related, associated*). Analyses were also run on a ‘lenient’ coding scheme to observe any differences due to coding decisions. This related specifically to the third and fourth coded questions. Ambiguous responses of “No” and “None” were coded as 0 for causal and 1 for correlational under this ‘lenient’ scheme.

The fact-recall questions, manipulation check and statistical knowledge check were all scored as 1 for every correct answer and 0 for every incorrect answer. A maximum fact-recall score was 10. As the manipulation check was aimed at the correction condition, a participant in the no-correction condition answering “I don’t know” for the manipulation check question was scored as 1. This is because it would not be reasonable to expect participants in this condition to comment on information they did not receive.

Data Preparation and Analytic Strategy

Four new variables were computed in SPSS for analyses: sum of causal inferences and sum of correlational inferences (each out of 5), sum of fact-recall answers (out of 10) and mean trust in sciences and scientists scores. Bayesian analyses were run in JASP to test the primary hypothesis. Bayesian techniques were used, rather than frequentist, as they enable the quantification of the strength of evidence in favour of both the null and alternative hypotheses (Dienes & Mclatchie, 2018; Wagenmakers et al., 2018). Moderated regressions were run in SPSS using PROCESS macro v3, Model 1 for moderation (Hayes, 2018) to test the secondary hypotheses relating to the moderating role of trust in science and scientists and TV watching time on the debunking effect.

Deviations from the preregistration

In the preregistration we stated that participants would be excluded for failing to successfully answer the manipulation-check question. This was a mistake based on a misreading of Ecker, Lewandowsky, Swire et al. (2011) and an unjustified deviation from established practices. Moreover, it was realised that it would only affect the correction condition, unbalancing the groups and undermining the randomisation. Consequently, no participants were excluded from analysis under this criterion, in keeping with previous studies (Ecker, Lewandowsky, Swire et al., 2011).

Results

We specified in our pre-registration that participants who had a fact-recall score more than three standard deviations below the mean would be removed before analyses. Fact-recall scores across conditions were M = 7.29 out of a possible score of 10 (SD = 1.93). No participants from the final sample were excluded under this criterion.

Primary Hypotheses

Bayesian one-sided independent samples *t*-tests compared the means of the no-correction group (n = 66) and the correction group (n = 63) on each dependent variable. As stated in the introduction, our one-sided, preregistered hypothesis was that participants in the correction condition would make fewer causal inferences and more correlational inferences than participants in the no-correction condition. Given that this study was not a replication, the prior could not be informed by previous research. In the absence of a clear alternative, a default half-Cauchy prior was used for analyses (van Doorn et al., 2021). The reported Bayes factors indicate the relative likelihood of each respective directional hypothesis, as opposed to the null hypothesis, in light of these data (Wagenmakers et al., 2018). For example, BF_{10} = 10 would imply that the alternative hypothesis is ten times more likely than the null, given the evidence. As a rough guide to interpreting Bayes factors, Kass and Raftery (1995) suggest that BF_{10} = 3 to 20 can be considered positive evidence for the alternative hypothesis, 20 to 150 strong evidence, and > 150 very strong. Delta (δ) calculated using a default two-sided Cauchy prior per the recommendations of van Doorn et al. (2021) is a standardised effect size, which represents the population estimate of Cohen's *d*. The accompanying Bayesian credible interval (BCI) provides a range of values in which we can be 95% certain the true population effect size lies (van Doorn et al., 2021).

Group differences on sum of causal inferences revealed very strong evidence for the debunking effect. On average, participants in the no-correction condition made 1.5 causal inferences (SD = 1.18), compared to 0.78 in the correction condition (SD = 0.91). Based on these data, our directional hypothesis is 279 times more likely than the null (BF_{10} = 279.35). The size of the effect could be characterised as 'medium to large', δ = 0.64, 95% BCI [0.28, 0.99]. When conceptualised as a probability of superiority, this effect indicates that there was a 68% chance that a person selected from the correction group would make fewer references to the misinformation than a participant from the no-correction group, if both were selected at random.

The analyses for sum of correlational inferences also showed very strong evidence for the alternative hypothesis over the null. On average, participants in the no-correction condition made 0.18 correlational inferences (*SD* = 0.43), whereas those in the correction group made 1.05 (*SD* = 0.83). The observation of such differences is 1,360 million times more likely under our hypothesis than the null, BF_{10} = 1.36e+09. This effect was substantial, $\delta = -1.27$, 95% BCI [-1.66, -0.90]. There was an 82% chance that a random participant in the correction condition would make more correlational inferences than a random participant in the no-correction condition.

Exploratory analyses were performed on participants who passed the manipulation-check ($n = 88$), on those who passed the statistical knowledge-check ($n = 85$) and on the 'lenient' coding scheme ($n = 129$). All showed a similar pattern of results, and conclusions remained unchanged. Furthermore, robustness analyses indicated that these conclusions were invariant across a range of reasonable prior widths (van Doorn et al., 2021). Finally, frequentist parallels of these *t*-tests were all statistically significant ($p < .005$). In other words, the frequentist and Bayesian analyses lead to the same conclusions.

Secondary Hypotheses

Exploratory analyses tested whether the debunking effect was moderated by trust in science and scientists, or the amount of time people watch TV. The attention of the reader is directed toward the interaction effects as these relate specifically to the test of our hypotheses. Descriptive statistics are presented in Table 1, which show that the groups did not markedly differ on these variables. Standardised Residuals, Mahalanobis and Cook's

Distances were calculated on the data and revealed no influential outliers for trust in science and scientists. However, seven outliers were identified for TV watching time. These cases were removed before analyses relating to TV watching time.

Table 1

Descriptive statistics for trust in science and scientists and TV watching time by condition

	Trust in Science and Scientists		TV Watching Time (hours/day)	
Condition	*N*	*M (SD)*	*N*	*M (SD)*
No-Correction	66	3.65 (0.41)	61	2.57 (1.82)
Correction	63	3.71 (0.52)	61	2.1 (1.56)

Moderated regressions using mean-centred variables were run separately for trust in science and scientists and TV watching time on each outcome variable (Hayes, 2018; Iacobucci et al., 2017). Condition was included as a predictor in both sets of moderated regressions. Analyses revealed significant main effects of condition on sum of causal inferences (b = -.72, p < .001) and sum of correlational inferences (b = .83, p < .001). Interaction effects between condition and trust in science and scientists were nonsignificant for sum of causal inferences (b = -.25, p = .60) and sum of correlational inferences (b = .37, p = .20).

For TV watching time, analyses revealed significant main effects of condition on both dependent variables: sum of causal inferences (b = -.66, p < .05), sum of correlational inferences (b = .88, p < .001). However, the interaction effects between condition and TV watching time were also non-significant for both dependent variables: sum of causal inferences (b = -.4, p = .70), and sum of correlational inferences (b = -.01, p = .94). In summary, these non-significant interaction effects do not support either of our secondary hypotheses. Neither trust in science and scientists, nor TV watching time influenced the effect of the correction.

Discussion

The results confirmed our primary hypotheses. Not only did the correction message significantly reduce the number of references to the misinformation, but it was also itself remembered and recalled by many participants. The medium-to-large effect size observed for sum of causal inferences is in line with the medium-to-large effect sizes reported in two meta-analyses of similar studies (Chan et al. 2017, Walter & Murphy, 2018). Walter and Murphy (2018) showed that debunking effect sizes were generally lower in online settings, when correcting real-world instead of constructed misinformation, and when using lay participants rather than students. This may explain why our effect was not as large as some that have been previously reported. The pattern of results for sum of correlational inferences also corresponds to descriptive data reported by previous research, showing that the correction group made more references to the correction (Ecker, Lewandowsky & Apai, 2011). These results are best accounted for by the mental models theory as the causal explanation provided in the correction allowed participants to fully update their mental model of the story without leaving a gap in their episodic memory (Lewandowsky et al., 2012). However, the secondary hypotheses were not supported by these data. Trust in science and scientists and TV watching time did not significantly influence participants' references to either the misinformation or the correction in these data.

Comparing this study with previous research, the correction reduced references to the misinformation by around a half, something which has been repeatedly noted by other authors (Ecker et al., 2017; Lewandowsky et al., 2012). Although a baseline group with which to measure the exact level of misinformation persistence could not be included (because it would not be possible to talk about a result without, at least tacitly, implying either correlation or causation), the large number of causal inferences still made by the correction group indicates that it was present to some degree. Conceivably, this may be heightened by the similarity and conceptual overlap between causal and correlational evidence, making the two easier to confuse. Still, the ubiquity of this finding within this field suggests something more fundamental about the way in which information is encoded and retrieved.

Our findings indicate that the traditional paradigm can be used to experimentally study the correction of statistical misinformation. We wanted to make correction as strong as possible, so we incorporated different debunking strategies, such as the correction coming from a credible source, providing a detailed alternative explanation which maintained coherence within the story, and explaining why misinformation was inaccurate within the correction (Lewandowsky et al., 2012). Prior research has tended to study these approaches individually, but we found that it is possible to successfully combine several of these methods into one correction.

To date, research in this field has been largely based around general misinformation (e.g., describing an event). This study expanded on what has been previously learnt by applying these principles and methods to a different type of misinformation: statistical terms. Although the examples of general misinformation traditionally used and the statistical misinformation examined here (causal versus correlational) are both text-based, the two types are quite different in nature and, as a result, translating the paradigm to this new context was challenging for the following reasons. First, it is easy to conflate correlational findings as causal, precisely because of the language used to describe them (Adams et al., 2017; Bleske-

Rechek, 2019). For example, *x leads to y* is a straightforward causal inference, whereas *x can lead to y* is more ambiguous (Adams et al., 2017). The latter could just as easily be saying that, while a correlation has been found, not all criteria have been met to establish cause and effect (Bleske-Rechek, 2019). This conceptual overlap meant that a large number of participants' answers were too ambiguous to be definitively coded, a problem not encountered in previous uses of the paradigm, which typically involves factual events. Second, considering that statistics are usually used as an aid, it was also difficult to only issue the misinformation in a single message, without implying it elsewhere to form the narrative of the story, and avoid increasing the potency of the misinformation through familiarity and coherence (Lewandowsky et al., 2012).

Despite explicitly asking participants to answer 'based on the story' to every coded inference question, a large number of answers reflected personal opinion instead. Several responses indicated that the participant inferred one thing from the story (e.g., correlation) but believed another (e.g., causal). This required additional thought when coding and calls into question how suitable inferences are for measuring corrections. It also reinforces the theory that the credibility of new information is processed in terms of its compatibility with the individual's pre-existing mental model and prior knowledge (Lewandowsky et al., 2012). The fact that many respondents neglected the instruction to answer 'based on the story' supports the idea that information processing is not a passive process, and provides support for accounts of motivated reasoning as responses reflected previously held beliefs rather than information supplied in the story (Kunda, 1990).

Although a large proportion of participants clearly interacted with the information, answering based on internal motives or reasoning, such reasoning did not appear to be influenced by trust in science and scientists or personal TV watching time. These findings differ from studies that have observed motivational effects, such as the worldview backfire effect, and instead support the assertion that these effects are not necessarily the norm (Swire et al., 2017). Ecker et al. (2014) suggested that these varied findings are explained by the degree to which attitude-incongruent corrections require an attitude change. The sample's relatively high levels of trust in science and low TV watching time might have meant that these variables did not require an attitude change and were not sufficiently emotive in this sample to influence the effect. These secondary analyses may also have been underpowered as interaction effects in moderated regressions tend to be small and can thus require very large sample sizes to detect (McClelland & Judd, 1993).

The non-significance of the moderation effects is still a worthwhile finding as we can be optimistic that the conditions around the debunking effect found here are not tied to these variables. For instance, it is not the case that people necessarily distrust science but rather they lack the tools to properly evaluate claims using scientific terms. This gives hope to the idea that simply increasing statistical literacy in the general population may be an important way to curb the spread of misinformation using statistical information. It also reaffirms the importance that reporters fully understand the statistical concepts they are reporting so as not to unintentionally misinform in the first instance.

Although just one type of statistical misinformation was investigated here, these findings have wider practical implications for how science is communicated to the general public, and how misinformation involving statistics can be corrected. Despite representing a well-educated sample of the public, only two thirds of the participants in this study were able to satisfactorily explain the difference between a causal and correlational link, as indicated by the statistical knowledge check. This finding highlights the ease with which this type of statistical information may be misrepresented and the importance of this research. Unlike previous studies in this field which have tended to use student samples, a lay sample was used here, making these results more generalisable and applicable to real-world situations where audiences may lack the critical evaluation skills and statistical background that students are likely to possess.

The results nonetheless showed that it is possible to successfully correct statistical misinformation: even though people may not have been able to *explain* the difference between correlation and causation, they were often able to *interpret* the data correctly once informed about its correlational nature. It follows that the media should carefully examine its use of statistical language, and correct it when necessary, to limit the spread of scientific misinformation. Furthermore, this study demonstrated that treatment of this type of misinformation may require additional thought relating to conceptual overlaps and coherency. The challenges in correcting causal misinformation, in addition to low levels of statistical literacy in the general population, suggests a greater emphasis on detailed alternative explanations over simple retractions is especially needed in this context. Although more difficult, it might be the case that correcting statistical misinformation has longer lasting benefits than correcting other types of misinformation, because explanatory corrections could improve statistical literacy (e.g., correlation does not imply causation), which can then be applied in other contexts. An informative correction may mean that people are able to take this knowledge forward when confronted with the same type of misinformation but relating to a different topic. Therefore, the misinformation may be less likely to take hold in the first place. Future research is needed to explore these possibilities.

Limitations and Future Research

Given the novel context, we kept the scope of the study modest to establish the presence of a debunking effect, and to understand how correcting statistical misinformation may differ to other types of information. Now that this has been achieved, future research should build on these findings, taking into account what has been learned about correcting misinformation in the context of statistics.

The lack of a distractor task could have impacted results due to the possible role of working memory (Thorson, 2016). This decision was taken given the online context to avoid making the study too laborious and leading to higher levels of attrition. It seems intuitive that a longer retention interval, in which attention is diverted to another task, might disrupt memory and increase the likelihood of misinformation persistence. However, Ecker, Lewandowsky and Apai (2011) observed debunking effects across similar experiments with retention intervals ranging from one to 40 minutes, although the sizes of these effects were not reported. In terms of longer delays, Walter and Murphy (2018) found in their meta-analysis that correction effects were stronger when beliefs were measured immediately following the correction than when a period of at least a day was allowed to pass before measuring the outcomes. However, whether statistical corrections are perhaps more salient and thus produce longer lasting effects is unknown. Future research could investigate whether a correction similar to that used here remained as effective after a short period of time (e.g. a distractor task) or a longer period of time (e.g. a separate testing session).

The online context of the study also meant that contextual unknowns may have been present during participation as a result of being unable to control environmental factors, which could have impacted the ability of some participants to remember relevant information when responding to questions (Gosling & Mason, 2015).

Although a sizable debunking effect was observed in the novel context of statistics, the boundary conditions surrounding this effect are unknown and would be a useful direction for future research. Given that statistics are so varied, future research could explore other common misuses of statistics and evaluate whether some types are easier to correct than others. For example, using graphs where the starting point of the y-axis has been manipulated may produce even larger debunking effects, as the reader would be visually confronted with how they have been misled.

Our data did not provide evidence of worldview consistency relating to trust in science and TV watching time as boundary conditions for the debunking effect. However, other possible boundary conditions would be worthy of exploration. For instance, the credibility of the source presenting the misinformation/correction or the effect of familiarity through repetition of the misinformation/correction (Walter & Tukachinsky, 2019). Such exploration may lead to an understanding of how the efficacy of corrections to statistical misinformation may be optimised or undermined.

Conclusion

This study was the first to investigate the correcting of statistical misinformation in the lay population, and whether an established paradigm used to correct general misinformation may be applied to statistics. These results provide strong evidential support for the ability to debunk misinformation about studies in which correlational evidence is misrepresented as causal, a key form of misinformation frequently presented to and consumed by the general public. Although we showed that the traditional paradigm can be successfully applied to debunking statistical misinformation, it is important to acknowledge the pitfalls relating to semantic ambiguity, as well as the challenge of measuring misinformation-persistence with this particular type of misinformation. The role of worldview consistency was not supported by these data, but warrants further research given the question around statistical power. By taking what has been learnt from previous research and applying it in a new context, this study has prompted several new lines of research in this area. These findings also have important practical implications for the correction of statistical misinformation, which is vital to limit the spread of misleading scientific findings in the media. As official bodies increasingly communicate information to the public through statistics, in turn influencing behaviour on important matters, research in this area has never been more essential. Future studies should seek to address ways in which the efficacy of corrections of statistical misinformation can be increased, and whether the findings here generalise to other types of statistical misinformation.

References

Adams, R., Sumner, P., Vivian-Griffiths, S., Barrington, A., Williams, A., Boivin, J., Chambers, C. & Bott, L. (2017). How readers understand causal and correlational expressions used in news headlines. *Journal of Experimental Psychology: Applied*, *23*(1), 1–14. https://doi.org/10.1037/xap0000100

Altay, S., & Mercier, H. (2020). Framing messages for vaccination supporters. *Journal of Experimental Psychology: Applied*, *26*(4), 567–578. https://doi.org/10.1037/xap0000271

Achterberg, P., de Koster, W., & van der Waal, J. (2017). A science confidence gap: Education, trust in scientific methods, and trust in scientific institutions in the United States, 2014. *Public Understanding of Science*, *26*(6), 704–720. https://doi.org/10.1177/0963662515617367

Bakalar, N. (2019, March 15). Can TV dumb you down? *The New York Times.* https://www.nytimes.com/2019/03/15/well/mind/tv-television-memory-brain-adults.html

Battley, P. (2019). *Kill or cure?* Kill or Cure. https://kill-or-cure.herokuapp.com/a-z/d

Berinsky, A. (2017). Rumors and health care reform: Experiments in political misinformation. *British Journal of Political Science*, *47*(2), 241-262. https://doi.org/10.1017/S0007123415000186

Bleske-Rechek, A. (2019, July 30). *The other crisis in psychology*. Quillette. https://quillette.com/2019/07/30/the-other-crisis-in-psychology/

Bleske-Rechek, A., Morrison, K., & Heidtke, L. (2015). Causal inference from descriptions of experimental and non-experimental research: Public understanding of correlation-versus-causation. *Journal of General Psychology*, *142*(1), 48–70. https://doi.org/10.1080/00221309.2014.977216

Chan, M., Jones, C., 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

Cook, J., & Lewandowsky, S. (2016). Rational irrationality: Modeling climate change belief polarization using Bayesian networks. *Topics in Cognitive Science*, *8*(1), 160–179. https://doi.org/10.1111/tops.12186

Cook, J., Lewandowsky, S., & Ecker, U. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PLoS ONE*, *12*(5), e0175799. https://doi.org/10.1371/journal.pone.0175799

Delaney, C. (2005). The spirituality scale: Development and psychometric testing of a holistic instrument to assess the human spiritual dimension. *Journal of Holistic Nursing*, *23*(2), 145–167. https://doi.org/10.1177/0898010105276180

Dienes, Z., & Mclatchie, N. (2018). Four reasons to prefer Bayesian analyses over significance testing. *Psychonomic Bulletin & Review*, *25*, 207-218. https://doi.org/10.3758/s13423-017-1266-z

Ecker, U., & Ang, L. (2019). Political attitudes and the processing of misinformation corrections. *Political Psychology*, *40*(2), 241–260. https://doi.org/10.1111/pops.12494

Ecker, U., Hogan, J., & Lewandowsky, S. (2017). Reminders and repetition of misinformation: Helping or hindering its retraction? *Journal of Applied Research in Memory and Cognition*, *6*(2), 185–192. https://doi.org/10.1016/j.jarmac.2017.01.014

Ecker, U., Lewandowsky, S., & Apai, J. (2011). Terrorists brought down the plane!— No, actually it was a technical fault: Processing corrections of emotive information. *The Quarterly Journal of Experimental Psychology*, *64*(2), 283–310. https://doi.org/10.1080/17470218.2010.497927

Ecker, U., Lewandowsky, S., Cheung, C., & Maybery, M. (2015). He did it! She did it! No, she did not! Multiple causal explanations and the continued influence of misinformation. *Journal of Memory and Language*, *85*, 101–115. https://doi.org/10.1016/j.jml.2015.09.002

Ecker, U., Lewandowsky, S., Fenton, O., & Martin, K. (2014). Do people keep believing because they want to? Preexisting attitudes and the continued influence of misinformation. *Memory & Cognition*, *42*(2), 292–304. https://doi.org/10.3758/s13421-013-0358-x

Ecker, U., Lewandowsky, S., Swire, B., & Chang, D. (2011). Correcting false information in memory: Manipulating the strength of misinformation encoding and its retraction. *Psychonomic Bulletin & Review*, *18*(3), 570–578. https://doi.org/10.3758/s13423-011-0065-1

Ecker, U., Lewandowsky, S., & Tang, D. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition*, *38*(8), 1087–1100. https://doi.org/10.3758/MC.38.8.1087

Flynn, D., Nyhan, B., & Reifler, J. (2017). The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics. *Political Psychology, 38*, 127-150. https://doi.org/10.1111/pops.12394

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Gosling, S. D., & Mason, W. (2015). Internet research in psychology. *Annual Review of*

Psychology, 66, 877-902. doi:10.1146/annurev-psych-010814-015321

Hart, P., & Nisbet, E. (2012). Boomerang effects in science communication: How motivated reasoning and identity cues amplify opinion polarization about climate mitigation policies. *Communication Research*, *39*(6), 701–723. https://doi.org/10.1177/0093650211416646

Hayes, A. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed). Guildford Press.

Hyman, I., & Jalbert, M. (2017). Misinformation and worldviews in the post-truth information age: Commentary on Lewandowsky, Ecker, and Cook. *Journal of Applied Research in Memory and Cognition*, *6*(4), 377–381. https://doi.org/10.1016/j.jarmac.2017.09.009

Iacobucci, D., Schneider, M. J., Popovich, D. L., & Bakamitsos, G. A. (2017). Mean centering, multicollinearity, and moderators in multiple regression: The reconciliation redux. *Behavior Research Methods*, *49*(1), 403–404. https://doi.org/10.3758/s13428-016-0827-9

Johnson, H., & Seifert, C. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(6), 1420–1436. https://doi.org/10.1037/0278-7393.20.6.1420

Kass, R., & Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773. https://doi.org/10.1080/01621459.1995.10476572

Ksir, C., & Hart, C. (2016). Correlation still does not imply causation. *The Lancet Psychiatry*, *3*(5), 401. https://doi.org/10.1016/s2215-0366(16)30005-0

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480

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

Lewandowsky, S., Ecker, U., Seifert, C., 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., Stritzke, W. G. K., Freund, A. M., Oberauer, K., & Krueger, J. I. (2013). Misinformation, disinformation, and violent conflict: From Iraq and the “War on Terror” to future threats to peace. *American Psychologist*, *68*(7), 487–501. https://doi.org/10.1037/a0034515

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

Matta, G. (2020). Science communication as a preventative tool in the COVID19 pandemic. *Humanities & Social Sciences Communications*, *7*(1), 1-14. https://doi.org/10.1057/s41599-020-00645-1

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, *114*(2), 376–390. https://doi.org/10.1037/0033-2909.114.2.376

Nadelson, L., Jorcyk, C., Yang, D., Jarratt Smith, M., Matson, S., Cornell, K., & Husting, V. (2014). I just don’t trust them: the development and validation of an assessment instrument to measure trust in science and scientists. *School Science and Mathematics*, *114*(2), 76–86. https://doi.org/10.1111/ssm.12051

Nyhan, B., & Reifler, J. (2015). Displacing misinformation about events: An experimental test of causal corrections. *Journal of Experimental Political Science, 2*(1), 81–93. https://doi.org/10.1017/XPS.2014.22

Pluviano, S., Watt, C., & Della Sala, S. (2017). Misinformation lingers in memory: Failure of three pro-vaccination strategies. *PLoS ONE*, *12*(7), e0181640. https://doi.org/10.1371/journal.pone.0181640

Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A., Recchia, G., van der Bles, A. M., & van der Linden, S. (2020). Susceptibility to misinformation about COVID-19 around the world. *Royal Society Open Science*, *7*(10), 201199. https://doi.org/10.1098/rsos.201199

Sallam, M., Dababseh, D., Eid, H., Al-Mahzoum, K., Al-Haidar, A., Taim, D., Yaseen, A., Ababneh, N. A., Bakri, F. G., & Mahafzah, A. (2021). High Rates of COVID-19 Vaccine Hesitancy and Its Association with Conspiracy Beliefs: A Study in Jordan and Kuwait among Other Arab Countries. *Vaccines*, *9*(1), 42. https://doi.org/10.3390/vaccines9010042

Scheufele, D., & Krause, N. (2019). Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences*, *116*(16), 7662–7669. https://doi.org/10.1073/pnas.1805871115

Sinatra, G., Kienhues, D., & Hofer, B. (2014). Addressing challenges to public understanding of science: Epistemic cognition, motivated reasoning, and conceptual change. *Educational Psychologist*, *49*(2), 123–138. https://doi.org/10.1080/00461520.2014.916216

Swire, B., Berinsky, A. J., Lewandowsky, S., & Ecker, U. K. H. (2017). Processing political misinformation: Comprehending the Trump phenomenon. *Royal Society Open Science*, *4*(3), 160802. https://doi.org/10.1098/rsos.160802

Szucs, D., & Ioannidis, J. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLoS biology*, *15*(3), e2000797. https://doi.org/10.1371/journal.pbio.2000797

Sumner, P., Vivian-Griffiths, S., Boivin, J., Williams, A., Venetis, C. A., Davies, A., Ogden, J., Whelan, L., Hughes, B., Dalton, B., Boy, F. & Chambers, C. D. (2014). The association between exaggeration in health related science news and academic press releases: retrospective observational study. *BMJ, 349*. https://doi.org/10.1136/bmj.g7015

Thorson, E. (2016). Belief Echoes: The Persistent Effects of Corrected Misinformation, *Political Communication, 33*(*3*), 460-480, https://doi.org/10.1080/10584609.2015.1102187

van der Linden, S. van der Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global Challenges*, *1*(2), 1600008. https://doi.org/10.1002/gch2.201600008

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š., Ly, A., Marsman, M., Matzke, D., Gupta, A., Sarafoglou, A., Stefan, A., Voelkel, J. G., & Wagenmakers, E. J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review, 28*(3) 813-826. https://doi.org/10.3758/s13423-020-01798-5

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*(1), 35–57. https://doi.org/10.3758/s13423-017-1343-3

Walter, N., & Murphy, S. (2018). How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs*, *85*(3), 423–441. https://doi.org/10.1080/03637751.2018.1467564

Walter, N., & Tukachinsky, R. (2019). 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

Wilkes, A., & Leatherbarrow, M. (1988). Editing episodic memory following the identification of error. *The Quarterly Journal of Experimental Psychology*, *40*(2), 361–387. https://doi.org/10.1080/02724988843000168

Wood, T., & Porter, E. (2019). The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behavior*, *41*(1), 135–163. https://doi.org/10.1007/s11109-018-9443-y