Correcting vaccine misinformation: A failure to replicate familiarity or fear-driven backfire effects

Ullrich K. H. Ecker[1,2*], Caitlin X. M. Sharkey[1] & Briony Swire-Thompson[3,4]

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

[2] Public Policy Institute, University of Western Australia, Perth, Australia

[3] Network Science Institute, Northeastern University, Boston, USA

[4] Institute of Quantitative Social Science, Harvard University, Cambridge, USA

* Corresponding author

E-mail: ullrich.ecker@uwa.edu.au

Abstract

Individuals often continue to rely on misinformation in their reasoning and decision making even after it has been corrected. This is known as the continued influence effect, and one of its presumed drivers is misinformation familiarity. As continued influence can promote misguided or unsafe behaviours, it is important to find ways to minimize the effect by designing more effective corrections. It has been argued that correction effectiveness is reduced if the correction repeats the to-be-debunked misinformation, thereby boosting its familiarity. Some have even suggested that this familiarity boost may cause a correction to inadvertently *increase* subsequent misinformation reliance; a phenomenon termed the familiarity backfire effect. A study by Pluviano et al. (2017) found evidence for this phenomenon using vaccine-related stimuli. The authors found that repeating vaccine "myths" and contrasting them with corresponding facts backfired relative to a control condition, ironically increasing false vaccine beliefs. The present study sought to replicate and extend this study. We included four conditions from the original Pluviano et al. study: the myths vs. facts, a visual infographic, a fear appeal, and a control condition. The present study also added a "myths-only" condition, which simply repeated false claims and labelled them as false; theoretically, this condition should be most likely to produce familiarity backfire. Participants received vaccine-myth corrections and were tested immediately post-correction, and again after a seven-day delay. We found that the myths vs. facts condition reduced vaccine misconceptions. None of the conditions increased vaccine misconceptions relative to control at either timepoint, or relative to a pre-intervention baseline; thus, no backfire effects were observed. This failure to replicate adds to the mounting evidence against familiarity backfire effects and has implications for vaccination communications and the design of debunking interventions.

Introduction

Misinformation—used here as an umbrella term for any information that is objectively false—has been identified as a serious issue for contemporary societies. This is not least because beliefs formed from invalid information may lead to behaviours that are potentially harmful or undesirable [1,2,3]. Vaccine misinformation is a prime example of this, as illustrated by the negative effect of false information about the mumps-measles-rubella (MMR) vaccine, or more recently the COVID-19 vaccines, on uptake rates [4,5]. One of the insidious characteristics of misinformation is that it can continue to influence people's reasoning and decision making even after it has been credibly corrected, a phenomenon known as the continued influence effect [6,7,8]. The persistent nature of misinformation has attracted much research seeking to examine the best methods to debunk "myths" in a way that most effectively reduces their subsequent impact (note we use the term "myth" to refer to a piece of common real-world misinformation) [9]. To this end, the present study sought to replicate a vaccine misinformation study by Pluviano et al. (2017) [10], who reported a failure of three debunking strategies.

Formatted: Font: 12 pt

It is generally acknowledged that the continued influence effect is at least partially based on failures of memory updating and retrieval processes [8]. One specific theoretical account—drawing on dual-process theories of memory [11]—posits that reliance on corrected misinformation occurs when a cue triggers retrieval of the misinformation based on its familiarity, but without recollection of the corresponding correction [12]. The familiarity of a myth has therefore been suggested as a driver of continued influence [13]. It is also well-known that repetition of information makes it more familiar and thereby more believable. This phenomenon is known as the illusory truth effect [14,15,16]. This effect occurs whether the information is true or false, and even if information conflicts with existing, factual knowledge [17,18]. Concern over illusory truth effects has led to the assumption that

repeating misinformation within a correction may render the correction less effective by boosting the familiarity of the misinformation being corrected. Some have even argued that corrections can backfire due to the boost to the misinformation's familiarity, and ironically increase the very misconception they are designed to reduce, relative to either a pre-correction baseline or a no-correction control group [19,20]. Demonstrations of such backfire effects have led to recommendations to avoid misinformation repetition when debunking misinformation [19,21,22].

However, as reviewed in detail elsewhere [8,23], the evidence for such familiarity backfire effects is actually quite weak: (1) The most cited study reporting familiarity backfire [Skurnik et al., 2007 [unpublished]; summarized in 20] is not accessible as a preprint. (2) Many studies claiming to have found familiarity backfire in fact only demonstrate a to-be-expected belief regression post-correction (i.e., a correction initially reduces belief and this corrective effect slowly wears off over time, with belief returning back to baseline; [24,25, but see 26 for a failed replication]). (3) There is ample evidence that familiarity backfire effects do not emerge even under conditions designed to be maximally conducive [13,27,28,29]. For example, Swire et al. (2017) presented participants with real-world myths, and corrected them using either brief or detailed explanations, which resulted in each false claim being presented three times during the experiment (thus boosting claim familiarity). Swire et al. tested young and older adults and varied the study-test delay from minutes to three weeks—the rationale being that (i) older adults should be more susceptible to familiarity effects because their ability to recollect details of the correction should be impaired, whereas familiarity-based memory is relatively unaffected by age [30], and that (ii) substantial delays should promote familiarity effects because recollection is affected more strongly by delays than familiarity [11]. However, corrections *reduced* belief in false claims in all conditions—even when the corrections were scant on detail, in older adults, and after a

three-week delay. Some have argued that familiarity backfire effects are mainly a concern with novel misinformation, because a correction may then introduce a person to a false claim they have never encountered before (thus providing a maximal familiarity boost, so to speak) [22]. However, evidence for this is also mixed at best [31,32,33].

One of the best pieces of evidence for familiarity backfire effects is a study by Pluviano et al. (2017) [10]. Pluviano and colleagues investigated how corrections of childhood vaccine myths impacted (i) concerns about vaccine side effects, (ii) belief in the debunked link between the MMR vaccine and autism, as well as (iii) vaccination intention (vaccine hesitancy). The study randomly assigned participants to one of four conditions: a common vaccine "myths-versus-facts" condition; a visual-correction condition utilising an infographic comparing disease and vaccine risks; a fear-appeal condition using images of sick (unvaccinated) children; or a control condition presenting unrelated fact sheets about healthcare. Participants' vaccine-myth beliefs and vaccination intentions were measured immediately after receiving a correction intervention (Time 1; T1), and again after a one-week delay (Time 2; T2), using the same questionnaire (note that a baseline measure using a more generic questionnaire was obtained before the intervention [Time 0; T0]; however, this measure was not included in any of Pluviano et al.'s analyses). None of the interventions substantially reduced misconceptions concerning vaccines relative to control. Instead, the myths-vs.-facts condition appeared to increase participants' belief in vaccine side effects and the vaccine-autism link, as well as reducing their intention to vaccinate, at T2. In a subsequent study, Pluviano et al. (2019) replicated this pattern of results in a parent population [34], where participants in a myths-vs.-facts condition held stronger misconceptions regarding vaccine side effects and the vaccine-autism link compared to control. This can be interpreted as evidence for familiarity backfire because the myths-vs.-facts format explicitly repeated the misinformation, boosting its familiarity. However, it

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

should also be noted that the fear appeal in Pluviano et al. (2017) also backfired, although this may have been for different reasons, most likely a misattribution of emotional arousal (see below).

The selection of conditions in Pluviano et al. (2017) was well-considered from both applied and theoretical perspectives. First, the myths-vs.-facts format is a very common format to address misconceptions in the real world, for example via posters or pamphlets. It is also theoretically interesting: On the one hand, offering an explanation about why a myth is false is a key ingredient of an effective correction, particularly when also providing a factual alternative [6,9,29,35]. On the other hand, the format has the potential to backfire because it involves repetition of the misinformation.

Second, visual interventions such as infographics are also commonly used in attempts to correct misinformation. They promise persuasiveness through attention capture and engagement and effective communication of complex concepts (e.g., weight-of-evidence messages and scientific belief representations) [36,37] that leaves little room for misinterpretation and counterarguments [38,39,40].

Finally, fear appeals are sometimes used in misinformation interventions where there is a relevant and significant threat. In such cases, fear appeals have been shown to be effective as long as there is high self-efficacy (i.e., the recipient has a sense that they can actively do something to avert the threat, such as quit smoking) [41,42,43]. In addition, emotive images used in fear appeals may increase their overall persuasiveness [44]. However, in line with Pluviano et al. (2017), Nyhan et al. (2014) [45] found that providing images of sick children depicting the symptoms of disease in a pro-vaccination campaign may be counterproductive, as misattribution of emotional arousal can potentially increase vaccine concerns [46].

The Present Study

Although the focus of our theoretical interest was on the familiarity backfire effect (and thus the comparison of myths-vs.-facts and control conditions), it was decided to include all of Pluviano et al.'s (2017) conditions. The present study thus replicated the Pluviano et al. (2017) study design, with several methodological enhancements. First, we additionally included a "myths-only" condition, which used the same materials as the myths-vs.-facts condition, but only presented the myths—labelled as such—without the facts. This format is an example of a weak, terse retraction that provides minimal correctional detail to recall later, and as such should be particularly likely to produce familiarity backfire [6,12]. Second, we included questions at baseline (T0) that were also given immediately post-correction (T1) and after a one-week delay (T2); this allowed for a within-subjects pre-post intervention comparison in addition to the between-subjects comparison between correction and no-correction conditions, to better establish the potential presence of a familiarity backfire effect. Third, we used multi-item measures instead of single-item measures to assess the dependent variables, because it is known that single-item measures often lack reliability, and their use has been causally related to observations of backfire [23,33].

A final change was motivated by the possibility that the backfire effect reported by Pluviano et al. (2017) was driven by worldview rather than familiarity. Such worldview backfire effects are occasionally observed when a correction challenges a misconception that a person is motivated to protect for ideological reasons [47,48]. These effects have also been difficult to replicate [49,50,51,52], but it is conceivable that worldview was an important factor in Pluviano et al.'s (2017) study. This is because worldview backfire effects have previously been found with vaccine stimuli [45,53] (but see [54]) and because Pluviano et al.'s sample was drawn from Italy and the UK, where vaccine hesitancy levels were relatively high at the time the study was conducted [55,56]. Thus, we added an empirically-tested scale

to assess vaccination attitudes, alongside a measure of identity centrality that assessed the importance of the vaccine attitude to the individual, as it has been suggested that only attitudes that are a central part of an individual's identity significantly impact reasoning [23,47]. Thus, a compound measure of vaccination attitudes and identity centrality was used as a covariate in the analyses, and also to allow for focused analysis of a subsample with relatively high vaccine concern, which may show worldview backfire effects.

Although Pluviano et al. (2017) found evidence for familiarity backfire effects, considering the overall body of research reviewed earlier, no backfire effect was expected. Therefore it was hypothesized that in the myths-only and myths-vs.-facts conditions, participants' beliefs in vaccine side effects, the vaccine-autism link, and vaccination hesitancy would be lower than control at both T1 and T2. It was also expected that there would be an initial decrease from T0 to T1 immediately post-correction, which, however, would not be fully sustained over time [13,23]; as such, it was expected that there would be an increase from T1 to T2. Regarding the other conditions, there was no reason to believe the visual correction would backfire, and thus it was expected that this condition would also be effective at reducing misconceptions. Finally, we had no strong expectations regarding the fear-appeal condition, given the inconsistent evidence from previous research, but again hypothesized that there would be no backfire. In sum, no experimental condition was predicted at T1 or T2 to exceed baseline levels at T0 or the control condition at T1 and T2, respectively.

Method

The core study design comprised the between-subjects factor condition with five levels (control; myths-only; myths vs. facts; visual correction; fear appeal) and the within-subjects factor time with two levels (immediate post-test, T1; delayed test, T2). Three dependent variables were measured (concern with vaccine side-effects; belief in the autism-

vaccine link; vaccination hesitancy) with seven items each. A subset of three items (one per dependent variable) was additionally administered at baseline (T0) to allow for a pre-post comparison. Vaccine attitudes and their identity centrality were measured at T0.

Participants

An a-priori power analysis using G*Power 3 [57] suggested a minimum sample size of 64 per condition to detect a difference of effect size $f = 0.25$ (with $\alpha = 0.05$; $1 - \beta = 0.80$) in between-subjects *F*-tests between the myths-vs.-facts and control conditions—the main comparisons of interest (note that the effect size was determined somewhat arbitrarily, but set to be smaller than the relatively large effect sizes reported by Pluviano et al. We also acknowledge that the *F*-tests referred to here are slightly different from the contrasts performed in the Results section (which were planned contrasts that take the full ANOVA model with all conditions into account and were subject to Holm-Bonferroni correction, which reduced achieved power). To ensure ample power and to account for exclusions (see below), it was decided to aim for 75 participants per condition, or a total sample size of 375. To additionally account for an expected drop-out rate of 15 % between T1 and T2, a convenience sample of 440 UK-based participants was recruited using Prolific. Of these, 383 participants completed both parts of the study. Based on a-priori exclusion criteria (see below), data of three participants were excluded, leaving a final sample of $N = 380$ (95 males, 283 females, 2 non-binary participants; mean age was $M = 36.45$ years [$SD = 11.66$], age range was 18-76). This sample size was large compared to Pluviano et al.' studies (2017, $N = 120$; 2019, $N = 60$). At the time of the study, 110 participants (29%) had a child under the age of six; this information was obtained to assess results in the parent population specifically, allowing for a comparison with Pluviano et al. (2019). Upon completion, participants received a compensation of £1 for Part 1 and £0.90 for Part 2.

Formatted: Font: 12 pt

Materials

Stimuli

Stimuli were taken directly from Pluviano et al. (2017) and are provided in the Supplement, available at https://osf.io/dwyma/.

Myths vs. Facts. Ten common vaccine misconceptions ("myths") were juxtaposed against 10 corresponding facts taken from World Health Organization educational materials (see Table S1). An example is "MYTH: Natural immunity is better than vaccine-acquired immunity. Indeed, catching a disease and then getting sick results in a stronger immunity to the disease than a vaccination." vs. "FACT: Vaccines interact with the immune system to produce a response similar to that produced by the natural infection, but they protect against its potential severe complications." Each myth/fact pair was presented on a separate page, with the fact appearing directly beneath the myth.

Myths Only. In this condition, only the 10 vaccine myths were presented, without the corresponding facts. Each was labelled explicitly as a myth and presented on an individual page. This condition was not part of the original Pluviano et al. (2017) study.

Visual Correction. In this condition, corrections visually compared the potential risk of symptoms if infected with a vaccine-preventable disease against the risk of vaccine side-effects. Individual diagrams for measles, mumps, and rubella were used, with each diagram showing 100 coloured stick figures to represent the degree of complications experienced—green (no/mild symptoms); yellow (moderate complications); and red (serious complications). This was supplemented by a short written explanation outlining the probability of experiencing these specific symptoms.

Fear Appeal. In this condition, participants were shown three photographs depicting unvaccinated children with symptoms of mumps, measles, and rubella. Images were accompanied by a personalized written warning stating that "you will see some of the

consequences you may face by choosing to not vaccinate your child". It also featured a series of dot points providing details about disease-specific infection risk and symptoms (e.g., "The measles virus can be spread very easily"; "Measles also can cause pneumonia, brain damage, seizures or death").

Control. Two fact sheets unrelated to vaccination safety were used in the control condition. One sheet contained 20 tips on how to prevent medical errors, while the other outlined five steps to safer healthcare. Participants viewed both fact sheets.

Measures

Pre-Manipulation Survey. The pre-manipulation survey administered at T0 contained three items assessing participants' baseline side-effect concerns, belief in a vaccine-autism link, and vaccine hesitancy. These items were: "I am concerned about serious adverse effects of vaccines"; "Some vaccines cause autism in healthy children"; and "Getting vaccines is a good way to protect my future child(ren) from disease". Participants responded on Likert scales ranging from 0 – 5 (*strongly disagree – strongly agree*). Pluviano et al. (2017) also administered a pre-manipulation survey including these three items (plus five other items assessing general vaccine attitudes, which were assessed in the present study at T0 with the dedicated 12-item Vaccination Attitude Examination scale described below). Although Pluviano et al. did not report any results from the pre-manipulation survey, we included it because (i) it may have provided some framing or priming that potentially influenced results in Pluviano et al.'s study, and (ii) because in the present study, the baseline items were also included in the post-manipulation survey, allowing for a direct pre-post comparison between T0 and both T1 and T2.

Post-Manipulation Survey. The 21-item post-manipulation survey was more specific to the intervention materials; it used seven items each to assess (i) belief in side effects (two items reverse-coded), (ii) the vaccine-autism link (three items reverse-coded), and (iii)

vaccine hesitancy (four items reverse-coded), respectively. Three items were taken directly from the Pluviano et al. (2017) materials (one per measure: "How likely is it that children who get the measles, mumps, and rubella [MMR] vaccine will suffer serious side effects?"; "Some vaccines cause autism in healthy children."; and "How likely is it that you would give your future child(ren) the MMR vaccine?"). These items were presented first, to allow for a direct replication of the Pluviano et al. analyses; these items were supplemented by new additional items (six per measure) to increase reliability. Responses were recorded on Likert scales ranging from 0 – 5 (*strongly disagree – strongly agree* or *very unlikely – very likely*). The post-manipulation survey was administered twice—once immediately post-intervention at T1, and again after a one-week delay at T2.

Vaccination Attitude Examination (VAX Scale). The VAX scale [58] consists of twelve items assessing general vaccine attitudes, including mistrust of vaccine benefits, worries about unforeseen future effects, concerns about commercial profiteering, and a preference for natural immunity. An example item is "I feel safe after being vaccinated". Responses were recorded on Likert scales ranging from 0 – 5 (*strongly disagree – strongly agree*); three items were reverse-coded. The VAX scale has high internal consistency ($\alpha = .92$) and good convergent and construct validity [59].

Identity-Centrality Survey. To assess the importance of vaccine beliefs and attitudes to participants' identity, two items were administered. These items were "My views about vaccinations are central to my identity" and "Vaccinations are an important topic to me". Responses were measured on Likert scales ranging from 0 – 5 (*strongly disagree – strongly agree*). A participant's score on the identity centrality scale was multiplied by their VAX *z*-score; this compound measure was then *z*-transformed to create a "VAX-ID" vaccination-attitude score that was used as a covariate in the analyses.

Procedure

The experiment was approved by the Human Research Ethics Office of the University of Western Australia (RA/4/20/6423). It was conducted online in May/June 2021, and administered using Qualtrics survey software (Qualtrics, Provo, UT). Participants initially received an information sheet and provided informed consent by ticking a box in the online survey before the study commenced. The information provided explained that the study was unrelated to COVID-19. Participants then (T0) answered the demographic questions (age, gender, and whether they had any children under the age of six). This was followed by the VAX scale, which used a fixed question order, as well as the identity-centrality scale and pre-manipulation survey, both of which used a randomised question order. Participants were then randomly assigned to one of the five conditions (control; myths-only; myths vs. facts; visual correction; fear appeal). After being presented with the respective intervention materials, all participants completed the post-manipulation survey (T1). In the post-manipulation survey, the three items taken from Pluviano et al. (2017) were always presented first, followed by the 18 new items in a quasi-random order (note that to minimize the number of response-scale switches, the original item using an agree/disagree response scale was presented first, followed by the two original items using a likely/unlikely response scale; this was followed by the two new questions using a likely/unlikely response scale [in random order], and finally the 16 new items that used an agree/disagree scale [also in random order]). After a week (T2), participants were invited back to complete Phase 2 of the study, where they were presented with the post-manipulation survey again. Phase 2 was open for ~ 48 hours. All stimuli and survey questions were presented for set minimum times (approx. 150 ms per word) to ensure that participants spent an adequate amount of time engaging with the written materials and questions. At the conclusion of the study, participants were asked whether their data should be used or discarded due to lack of effort, and were then fully debriefed. The

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

debriefing explained to participants that they may have been exposed to vaccine misinformation and how this may affect them. They were also given the ten facts from the myths-vs.-facts condition and links to relevant World Health Organization and National Health Service web pages (see Supplement). The experiment took approximately 15 minutes to complete (10 minutes for Phase 1; 5 minutes for Phase 2).

Results

Data were excluded from analysis based on a-priori criteria. Specifically, we first screened for participants who indicated their data should be discarded due to lack of effort ($n = 0$) and those who showed uniform responding ($SD < 0.5$ across all rating-scale items; $n = 0$). Scores of reverse-coded items were then reversed, and data were screened for inconsistent responding; this was done by (i) computing separate means for reverse-coded and regular items for the VAX scale and each of the three dependent measures at each of the two time-points, (ii) computing a grand mean from the seven absolute differences between those means, and (iii) applying the outlier-labelling rule with a 2.2 multiplier [60] to identify outliers on that score ($n = 3$). We also assessed reliability and found that the VAX scale and each dependent-variable scale demonstrated very good internal consistency (all Cronbach's $\alpha \geq .87$).

Primary Analyses: Side Effects, Vaccine-Autism Link, Vaccination Hesitancy

We first present the primary analyses of the three dependent variables measured using our multi-item scales and including our "VAX-ID" covariate. Supplementary analyses taking into account vaccination attitudes and parental status, analyses focused only on Pluviano et al.'s (2017) original items (i.e., a direct replication), and pre-post analyses are presented in a later section.

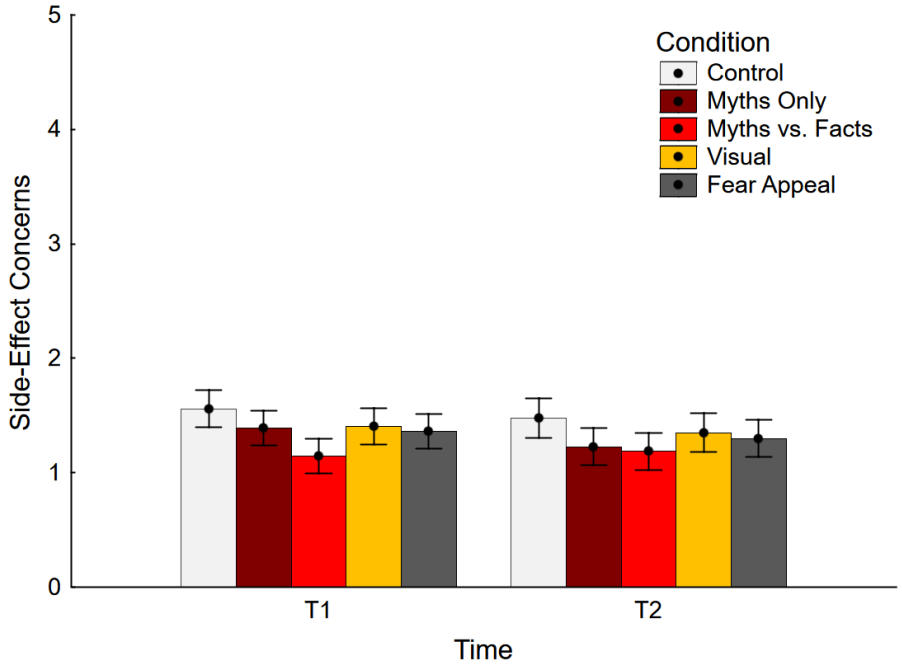
Three two-factorial within-between analyses of covariance (ANCOVAs) were conducted in order to examine whether beliefs in vaccine side effects, the vaccine-autism link, and vaccination hesitancy differed across time points and experimental conditions. The within-subjects factor time had two levels, T1 and T2; the between-subjects factor condition had five levels, reflecting the control, myths-only, myths-vs.-facts, visual-correction, and fear-appeal conditions. In order to take into account both participants' general vaccination attitudes and the identity centrality of those attitudes, the VAX-ID covariate was included in a full-factorial model (note that ANOVAs without the covariate yielded equivalent results unless noted otherwise).

Formatted: Font: 12 pt

Side-effect concern data are shown in Fig 1. The ANCOVA yielded a significant main effect of time, $F(1, 370) = 5.45$, $p = .020$, $\eta_p^2 = .015$, indicating slightly lower scores at T2 relative to T1. The main effect of condition was also significant, $F(4, 370) = 2.73$, $p = .029$, $\eta_p^2 = .029$, indicating a difference between test conditions (note that the effect was nonsignificant in an ANOVA, $F(4, 375) = 2.34$, $p = .055$, $\eta_p^2 = .024$). The interaction was not significant, $F(4, 370) = 1.44$, $p = .218$, $\eta_p^2 = .015$. As expected, the covariate had a significant impact, $F(1, 370) = 336.61$, $p < .001$, $\eta_p^2 = .476$, but was not involved in any interactions, all $F(1/4, 370) \leq 3.18$, $p \geq .075$, $\eta_p^2 \leq .009$.

Formatted: Font: 12 pt

Fig 1. Concerns About Side Effects Across Conditions. Error bars show 95% confidence intervals.



Planned contrasts were conducted to compare each experimental condition against control at each delay; these are presented in Table 1 (top section). The analyses revealed that the myths-vs.-facts condition was associated with reduced concern about vaccine side effects at Time 1 relative to control. In other words, this condition was effective at reducing side-effect concerns. No conditions were associated with elevated concerns relative to control at any time. This indicated that there was no backfire effect in any of the conditions.

Table 1. Contrast Analyses.

Contrast (vs. control)	$F(1, 370)$	η_p^2	p
Side Effects T1			
Myths Only	2.29	.006	.131
Myths vs. Facts	13.43	.035	< .001*
Visual	1.86	.005	.173
Fear Appeal	3.10	.008	.079
Side Effects T2			
Myths Only	4.24	.011	.040
Myths vs. Facts	5.71	.015	.017
Visual	1.02	.003	.313
Fear Appeal	2.11	.006	.147
Vaccine-Autism Link T1			
Myths Only	1.80	.005	.180
Myths vs. Facts	6.90	.018	.009*
Visual	0.13	< .001	.719
Fear Appeal	0.02	< .001	.884
Vaccine-Autism Link T2			
Myths Only	5.85	.016	.016
Myths vs. Facts	4.99	.013	.026
Visual	0.74	.002	.392
Fear Appeal	1.07	.003	.303
Vaccination Hesitancy T1			
Myths Only	0.03	< .001	.860
Myths vs. Facts	0.46	.001	.497
Visual	0.44	.001	.509
Fear Appeal	1.45	.004	.230
Vaccination Hesitancy T2			
Myths Only	1.75	.005	.186
Myths vs. Facts	0.99	.003	.320
Visual	0.59	.002	.443
Fear Appeal	0.10	< .001	.757

* indicates statistical significance following Holm-Bonferroni correction. Note that correction was applied to sets of contrasts defined by the combination of dependent variable and timepoint (i.e., family size 4), as per the a-priori analysis plan; however, one could also argue that correction should instead control only for the dual tests across timepoints (i.e.,

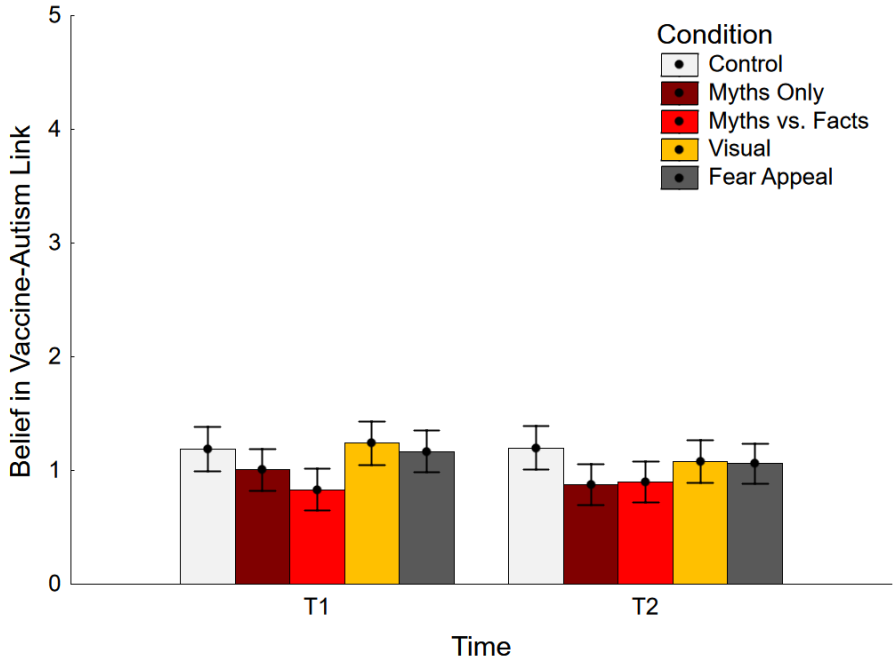
Formatted: Space After: 18 pt

Formatted: Font: 12 pt

family size 2), as only those test the same hypothesis (e.g., "myths-only differs from control"; see [61]). This would result in some non-significant contrasts becoming significant.

Data regarding belief in the vaccine-autism link are shown in Fig 2. The ANCOVA yielded a main effect of time, $F(1, 370) = 7.33$, $p = .007$, $\eta_p^2 = .019$, indicating lower scores at T2 relative to T1. The main effect of condition was significant as well, $F(4, 370) = 2.52$, $p = .041$, $\eta_p^2 = .027$ (note that the effect was nonsignificant in an ANOVA, $F(4, 375) = 2.27$, $p = .061$, $\eta_p^2 = .024$). There was also a significant interaction of condition and time, $F(4, 370) = 3.56$, $p = .007$, $\eta_p^2 = .037$. The covariate had a significant impact, $F(1, 370) = 264.72$, $p < .001$, $\eta_p^2 = .476$, but was not involved in any interactions, all $F(1/4, 370) \leq 1.60$, $p \geq .207$, $\eta_p^2 \leq .006$.

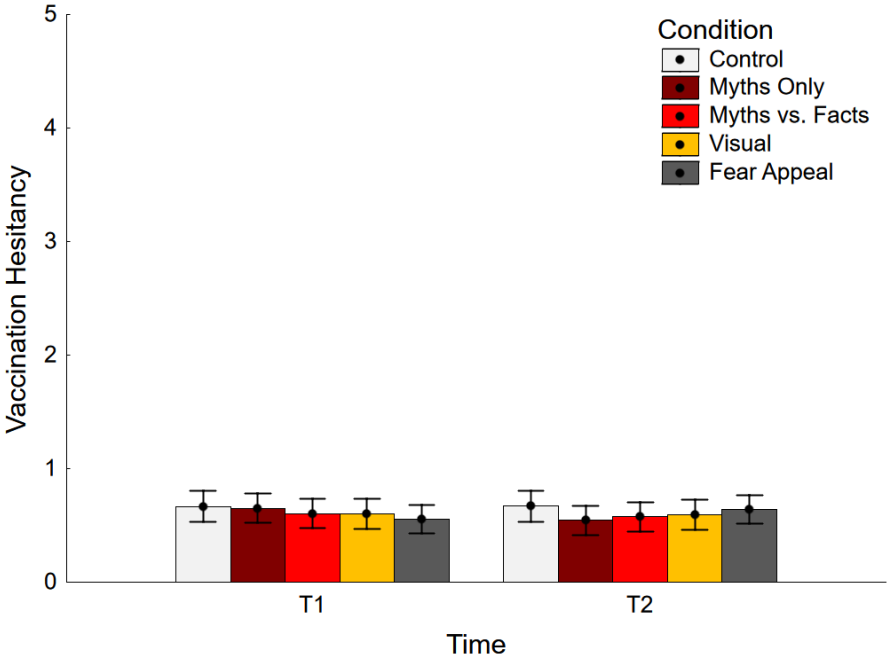
Fig 2. Belief in Vaccine-Autism Link Across Conditions. Error bars show 95% confidence intervals.



Planned contrast analyses showed that the myths-vs.-facts condition was associated with significantly lower belief in the vaccine-autism link at T1 (Table 1, middle section). No conditions were associated with a statistically significant belief increase relative to control at any time, indicating that there was no backfire effect present in any of the conditions.

Vaccination hesitancy results are presented in Fig 3. The ANCOVA returned non-significant main effects of time and condition, $F < 1$, but a significant interaction effect, $F(4, 370) = 3.51$, $p = .008$, $\eta_p^2 = .037$. However, no significant differences were found between control and the other experimental conditions, suggesting that no condition increased or decreased vaccine hesitancy, relative to control (refer to Table 1; bottom section). Again, this indicates that there was no backfire effect present in any condition.

Fig 3. Vaccination Hesitancy Across Conditions. Error bars show 95% confidence intervals.



Supplementary Analyses

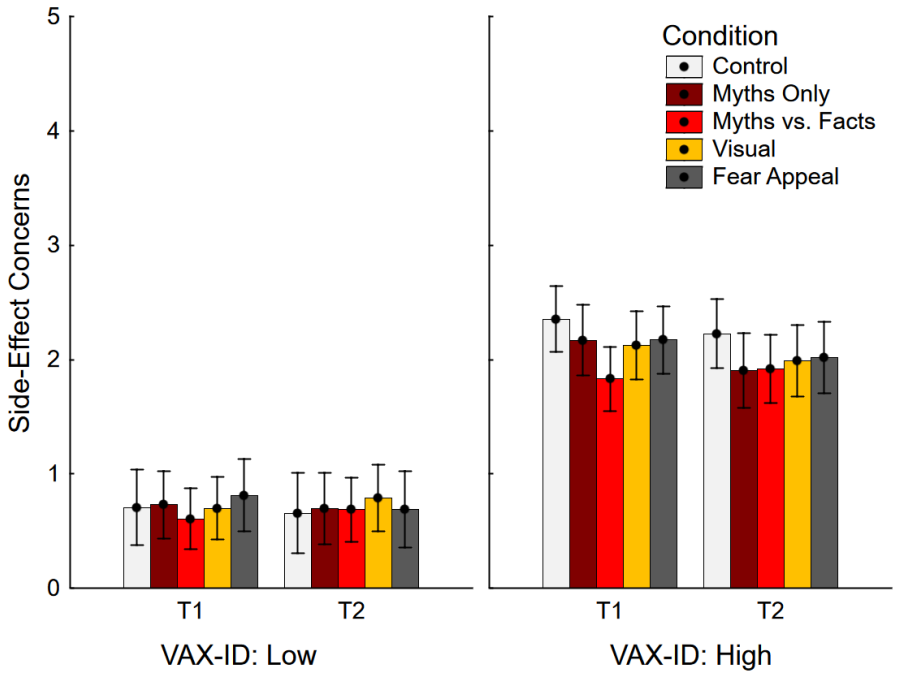
Analyses Considering Vaccination Attitudes

Given that the sample used in Pluviano et al. (2017) was drawn from a potentially more vaccine-hesitant population, supplementary analyses were performed on the top and bottom tertiles of the sample based on VAX-ID scores (n = 255). This involved repeated measures ANOVAs with the within-subjects factor time (T1, T2) and the between-subjects factors condition (control, myths only, myths vs. facts, visual, fear appeal) and VAX-ID group (top, bottom). As per a-priori analysis plan, this was followed by specific contrasts between control and experimental conditions in the top tertile regardless of ANOVA outcome, to ensure no potential backfire effect was missed. To foreshadow, no backfire effects emerged on any variable (note that exploratory analyses using more extreme groups [e.g., deciles] also found no evidence for backfire).

Formatted: Font: 12 pt

Concerns about side effects are shown in Fig 4. There was only a significant main effect of VAX-ID group, in the expected direction, $F(1, 245) = 221.28$, $p < .001$, $\eta_p^2 = .475$. The main effects of time, $F(1, 245) = 3.57$, $p = .060$, $\eta_p^2 = .014$, and condition, $F < 1$, were non-significant, and there were no significant interactions, all $F(1/4, 245) \leq 2.71$, $p \geq .101$, $\eta_p^2 \leq .028$. The planned contrast analysis focusing on the top tertile of vaccine-hesitant participants (see Table 2, top section) returned just one significant effect, suggesting reduced side-effect concerns in the myths-vs.-facts condition relative to control at T1.

Fig 4. Concerns About Side Effects Across Conditions and VAX-ID Groups. VAX-ID: product of vaccine-attitude and identity-centrality scores; T1: time 1; T2: time 2. Error bars show 95% confidence intervals.



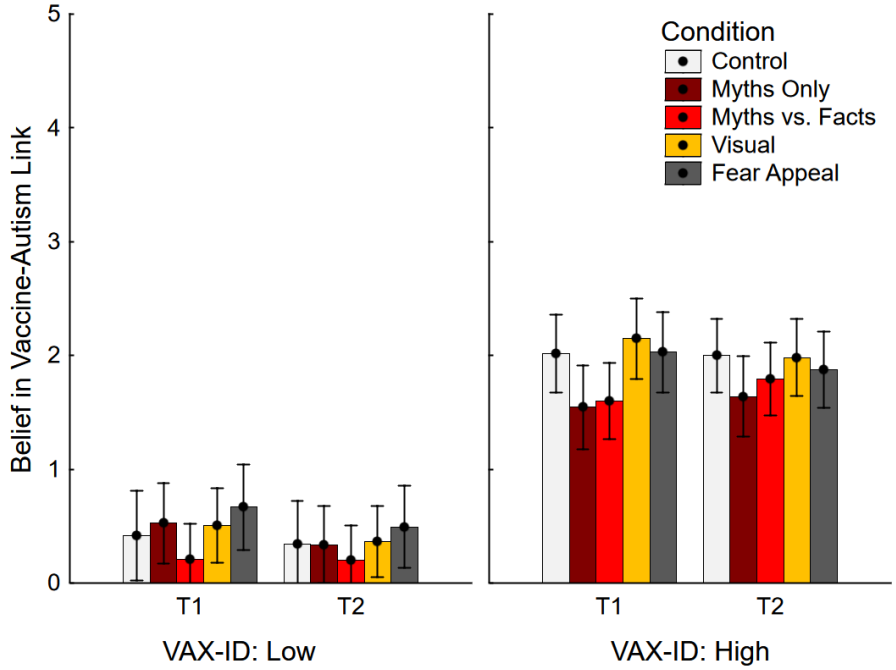
Data regarding belief in the vaccine-autism link is shown in Fig 5. The ANOVA returned significant main effects of time, $F(1, 245) = 5.17$, $p = .024$, $\eta_p^2 = .021$, indicating slightly lower scores at T2 than T1, and VAX-ID group $F(1, 245) = 185.37$, $p < .001$, $\eta_p^2 = .431$. The main effect of condition was non-significant, $F(4, 245) = 1.52$, $p = .197$, $\eta_p^2 = .024$, but there was a time by condition interaction, $F(4, 245) = 3.12$, $p = .016$, $\eta_p^2 = .048$. No other interactions were significant, all $F(1/4, 245) \leq 3.70$, $p \geq .056$, $\eta_p^2 \leq .024$. No contrasts were significant (see Table 2, middle section), indicating no significant impact of any interventions in the top tertile of vaccine-hesitant participants.

Table 2. Contrast Analyses in Vaccine-Hesitant Group.

Contrast (vs. control)	$F(1, 245)$	η_p^2	p
Side Effects T1			
Myths Only	0.77	.003	.382
Myths vs. Facts	6.65	.026	.010*
Visual	1.20	.005	.274
Fear Appeal	0.77	.003	.381
Side Effects T2			
Myths Only	2.03	.008	.156
Myths vs. Facts	2.08	.008	.150
Visual	1.18	.005	.279
Fear Appeal	0.91	.004	.341
Vaccine-Autism Link T1			
Myths Only	3.41	.014	.066
Myths vs. Facts	2.94	.012	.088
Visual	0.29	.001	.594
Fear Appeal	< 0.01	< .001	.959
Vaccine-Autism Link T2			
Myths Only	2.20	.009	.139
Myths vs. Facts	0.78	.003	.377
Visual	0.01	< .001	.943
Fear Appeal	0.28	.001	.597
Vaccination Hesitancy T1			
Myths Only	0.01	< .001	.910
Myths vs. Facts	0.28	.001	.599
Visual	0.16	.001	.691
Fear Appeal	0.52	.002	.473
Vaccination Hesitancy T2			
Myths Only	0.34	.001	.561
Myths vs. Facts	0.80	.003	.371
Visual	0.39	.002	.533
Fear Appeal	< 0.01	< .001	.949

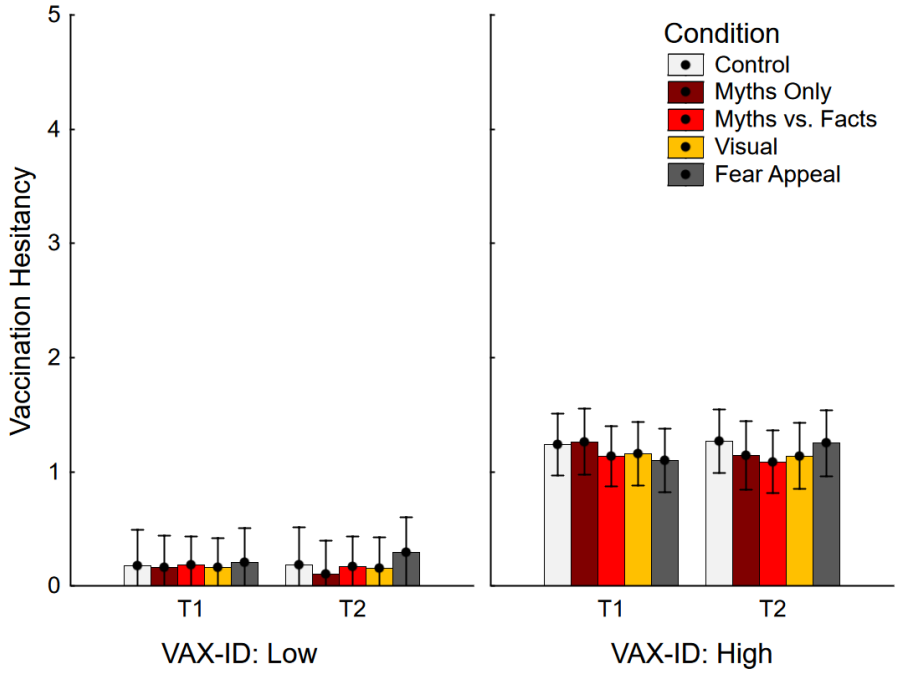
* indicates statistical significance following Holm-Bonferroni correction.

Fig 5. Belief in Vaccine-Autism Link Across Conditions and VAX-ID Groups. VAX-ID: product of vaccine-attitude and identity-centrality scores; T1: time 1; T2: time 2. Error bars show 95% confidence intervals.



Vaccine hesitancy data are shown in Fig 6. The ANOVA yielded the expected main effect of VAX-ID group $F(1, 245) = 126.92$, $p < .001$, $\eta_p^2 = .341$, but no other main effects, both $F < 1$. There was a marginal time by condition interaction, $F(4, 245) = 2.50$, $p = .043$, $\eta_p^2 = .039$, but no other significant interactions, all $F < 1$. No contrasts were significant (see Table 2, bottom section).

Fig 6. Vaccination Hesitancy Across Conditions and VAX-ID Groups. VAX-ID: product of vaccine-attitude and identity-centrality scores; T1: time 1; T2: time 2. Error bars show 95% confidence intervals.



Analyses Considering Parent Status

Next, analysis focused on participants with children, to allow comparison with Pluviano et al. (2019), who found a familiarity backfire effect in a parent sample. To this end, full-factorial repeated measures ANCOVAs were conducted with the within-subjects factor time (T1, T2), the between-subjects factors condition (control, myths only, myths vs. facts, visual, fear appeal) and parent status (yes, no), and VAX-ID as a covariate. (Note, this analysis was deemed appropriate despite unequal sample sizes, as ANOVA is relatively robust to sample size differences as long as variances are not also unequal. Nevertheless, analyses were repeated with equal sample sizes using a random subsample of non-parents; results were comparable.) There were no significant main effects or interactions involving

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

parent status across all three dependent variables, all $F(1/4, 360) \leq 2.15$, $p \geq .074$, $\eta_p^2 \leq .023$; this indicates that the effect of the experimental manipulations did not differ as a function of parent status. There were no backfire effects in any condition at any timepoint (see Table S2).

Replication Using Only Pluviano et al.'s (2017) Items

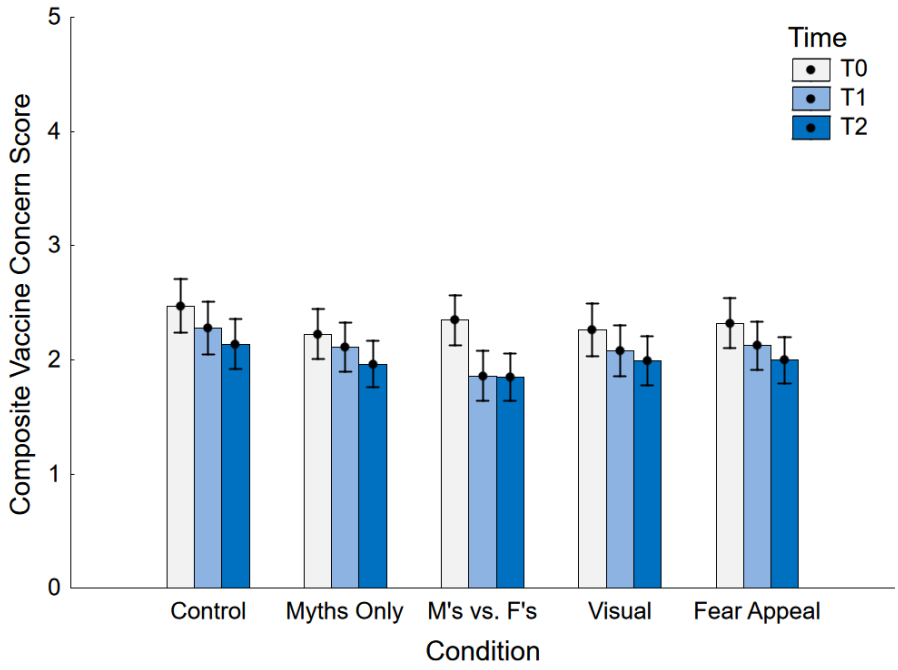
Next, we replicated the original Pluviano et al. (2017) analyses using their one-item measures; the myths-only condition and the VAX-ID covariate were dropped for these analyses as they were not part of the original design. Separate time (T1, T2) by condition (control, myths vs. facts, visual, fear appeal) ANOVAs on the three dependent measures yielded no significant main effects of condition, all $F(3, 297) \leq 1.38$, $p \geq .248$, $\eta_p^2 \leq .014$, and no time by condition interactions, all $F(3, 297) \leq 1.43$, $p \geq .234$, $\eta_p^2 \leq .014$. There were no backfire effects in any condition at any timepoint (see Table S3).

Pre-Post Analyses

Backfire effects are defined by an ironic increase in belief relative to either a control condition (as used in the primary analyses) or a pre-manipulation baseline (Swire-Thompson et al., 2020). We therefore conducted pre-post analyses, using a composite measure based on the three items administered at all time points (T0, T1, T2). Data are shown in Fig 7. First, a between-subjects ANOVA with the sole factor of condition (control, myths only, myths vs. facts, visual, fear appeal) was conducted at T0, to ascertain that there were no condition differences at baseline; no differences between conditions were found, $F < 1$. Then, a full-factorial repeated-measures ANCOVA with the within-subject factor time (T0, T1, T2), the between-subjects factor condition (control, myths only, myths vs. facts, visual, fear appeal), and the VAX-ID covariate was conducted. This yielded a significant main effect of time $F(2, 740) = 87.20$, $p < .001$, $\eta_p^2 = .191$, indicating a significant decrease over timepoints. There was the expected main effect of VAX-ID, $F(1, 370) = 518.51$, $p < .001$, $\eta_p^2 = .584$, but no main effect of condition, $F < 1$. However, a significant interaction between time and

condition was found, $F(8, 740) = 3.35$, $p = .001$, $\eta_p^2 = .035$, suggesting that the effect of the experimental manipulations varied across timepoints. There was also a time by VAX-ID interaction, $F(2, 740) = 10.28$, $p < .001$, $\eta_p^2 = .027$; closer inspection suggested this was due to stronger concern reduction over time in participants with greater VAX-ID scores (i.e., in those with greater vaccine concerns). There were no other significant effects, all $F < 1$.

Fig 7. Pre- and Post-Intervention Vaccine Concern Across Conditions. M's vs. F's: Myths vs. Facts; T0: time 0 (pre-intervention); T1: time 1; T2: time 2. Error bars show 95% confidence intervals.



Contrasts of T1 and T2, respectively, against T0 are presented in Table 3, separately for each condition. Aside from the myths-only condition at T1, all T1 and T2 scores were significantly lower than the T0 baseline, including the control condition. In line with the main analyses, when contrasted against control, only the myths-vs.-facts condition was associated

with lower concern at T1, $F(1, 375) = 12.61$, $p < .001$, $\eta_p^2 = .033$, but not T2, $F(1, 375) = 3.10$, $p = .079$, $\eta_p^2 = .008$.

Table 3. Pre-Post Contrast Analyses.

Contrast (vs. T0)	$F(1, 375)$	η_p^2	p
Control			
T1	10.70	.028	.001*
T2	23.52	.059	< .001*
Myths Only			
T1	4.11	.011	.043
T2	16.35	.042	< .001*
Myths vs. Facts			
T1	74.24	.165	< .001*
T2	58.97	.136	< .001*
Visual			
T1	9.41	.024	.002*
T2	15.70	.040	< .001*
Fear			
T1	12.67	.033	< .001*
T2	25.89	.065	< .001*

* indicates significance after Holm-Bonferroni correction

We finally examined the proportion of participants with numerically decreased misperceptions (corrective change), increased misperceptions (backfire), or no change post intervention. Proportions across conditions are summarized in Table 4.

Table 4. Proportions of Numerical Change Tendencies (in %) Across Conditions and Timepoints.

Numerical Change (from T0)	Corrective	No Change	Backfire
Control			
T1	54.29	25.71	20.00
T2	65.71	24.29	10.00
Myths Only			
T1	43.04	36.71	20.25
T2	55.70	27.85	16.46
Myths vs. Facts			
T1	65.38	29.49	5.13
T2	66.67	21.79	11.54
Visual			
T1	52.78	27.78	19.44
T2	58.33	23.61	18.06
Fear			
T1	55.56	29.63	14.81
T2	62.96	22.22	14.81

Discussion

Despite a growing number of studies finding evidence against familiarity backfire effects [13,26,27,28,29,32,33], several studies still claim familiarity to be a genuine mechanism for backfire effects [e.g., 22,31]; given the sound theoretical reasons to believe familiarity backfire effects can occur, more solid evidence is required. Moreover, the concept still creates concern amongst practitioners, and it is therefore important to scrutinize reports of the effect. The present study therefore replicated and extended a study by Pluviano et al. (2017), which found that fear appeals and corrections presented in a myths-vs.-facts format backfired, inadvertently increasing belief in vaccine misconceptions and vaccination hesitancy relative to a control condition [10].

Contrary to Pluviano et al.'s [10] findings, based on the overall literature, it was predicted that corrections would reduce—not strengthen—false beliefs in vaccine side effects and the MMR-vaccine-autism link, as well as vaccination hesitancy, relative to control. We also expected corrections to reduce misconceptions and hesitancy (at timepoint T1) relative to a pre-intervention baseline (timepoint T0), although we expected some potential belief regression over time (at timepoint T2 relative to T1). However, this regression was not expected to reach or exceed baseline levels pre-intervention (at T0).

Results largely confirmed these predictions. We found that no intervention was associated with greater misinformation belief or vaccine hesitancy than control at any timepoint. This was true across all analyses and subgroups, including in parents (at odds with [34]) and in those participants higher in anti-vaccination sentiment (broadly in line with [54]). In the following, we focus our discussion on the impact of the interventions on misconceptions, before we briefly address their impact on vaccine hesitancy.

Although no condition *increased* vaccine misconceptions, only the myths-vs.-facts condition successfully *decreased* belief in vaccine side effects and the vaccine-autism link relative to control. When comparing pre- and post-intervention misconceptions, all conditions but the myths-only condition were associated with reduced misconceptions post-intervention, including the control condition. Again, it was only the myths-vs.-facts condition that led to a significantly stronger reduction than control, without belief regression back to baseline after a week. We acknowledge that the study had limited power to detect small effects (e.g., some observed non-significant effects were in the range of $.01 < \eta_p^2 < .02$), so some interventions may have been found to be significantly effective with greater power. This should not, however, distract from our core finding that no intervention demonstrated any tendency of backfire.

One reason for the efficacy of the myths-vs.-facts condition in decreasing misconceptions may lie in the clear and detailed alternative information presented when refuting the myths. It has been suggested that provision of alternative, factual information is the most important ingredient of a successful correction, allowing individuals to update their understanding and replace false with factual information in their mental models of the world [6,8,9,35]. This also explains why no significant effect was found for the myths-only condition, which provided the weakest possible retraction [6,12]. Despite the myths-only condition theoretically being the one most likely to cause a familiarity-driven backfire effect, though, no such effect was observed; this is particularly strong evidence against the notion of familiarity backfire.

In regard to the visual correction, its lack of efficacy relative to control was unexpected. It can be speculated that participants may not have actively engaged with the infographics to the extent required in order to allow a proper risk evaluation. While graphically-provided information has been shown to be effective, and potentially superior to text alone [36,37,38,39,40], extracting meaning from graphical material still requires individuals' attention and engagement, even for low-level visual statistical learning [62]. The infographic used in the current study certainly did require attention to fully comprehend the colour coding and the relative-risk information conveyed. Infographics may thus only be useful for correcting misconceptions in situations where individuals are fully engaged with processing the information provided, or when the infographics are extremely simple. However, we again acknowledge that the study had limited power to detect small effects.

Finally, we did not find any evidence for a backfire effect in the fear-appeal condition either, which at the group level was also found ineffective relative to control. We note that this was not due to a pronounced bimodality (i.e., the intervention "working" for some participants but backfiring for others), as the backfire rate of approximately 15 % was

comparable to other conditions. The fact that intervention efficacy was relatively low overall, relative to control, is most likely associated with demand characteristics affecting responses in the control group. We note, however, that having a low level of misconception belief in the control condition will increase the likelihood of observing backfire effects. This is because the low belief in the control condition would leave sufficient leeway on the scale for the level of belief to surpass control in the other conditions.

Overall, the observed impact of the myths-vs.-facts condition is in line with recent evidence that has likewise found the format to be particularly efficacious, especially when compared to interventions that focus only on the facts without directly countering the myths [63,64]. However, alternative formats should not be neglected based on current findings. As Swire-Thompson et al. [63] discuss, the optimal format may depend on the specific content and context of the correction, and in general it is more important *that* a myth is corrected than what specific format is used (also see [65]). More work is required to ascertain the relative strengths of different formats, including visual corrections that have been shown to be effective in other contexts.

With regards to the interventions' impact on vaccine hesitancy, it was found that no intervention reduced vaccine hesitancy relative to control, even in participants with relatively high baseline levels of hesitancy. This is important because arguably the ultimate goal of any debunking intervention is to reduce misconceptions in order to change behavioural choices and outcomes. It is well-known that changes to beliefs and attitudes tend to not translate to equivalent changes in behavioural intentions and behaviours [66,67]. In fact, other research has found that misinformation corrections tend to have stronger impact on the targeted misconceptions than on related behaviours or behavioural intentions, including vaccination intentions [e.g., 2,45,68,69]. In the present study, it is possible that the observed effects are true small effects that would have been statistically significant with greater power and

potentially meaningful at scale. However, effect sizes were consistently smaller than $\eta_p^2 = .01$, and as such it is also possible that more than a brief one-off intervention is necessary to achieve any practically significant change in intentions and behaviours.

A remaining question is: Why did our findings differ from those of Pluviano et al. (2017) [10]? We offer several reasons. First, Pluviano et al. conducted their study in 2016, when skepticism towards childhood vaccines may have been somewhat greater than in 2021 [55]. The Pluviano et al. study also included participants from both the UK and Italy (whereas our participants were only from the UK), and it is possible that the Italian participants were particularly vaccine-skeptical [56]. The backfire effect observed by Pluviano et al. may have thus been driven by worldview rather than familiarity. This is perhaps even more likely given that Italy introduced mandatory childhood vaccinations in mid-2017 because of relatively low vaccination rates compared to other European countries [70]. This may not only highlight relatively greater vaccine skepticism in Italy (pre-2017), but also suggests that public discourse around the mandate may have polarized the Italian sample in Pluviano's study (we thank one of the reviewers, Dr Aimee Challenger, for pointing out this policy change in Italy). However, it is important to note that we did not observe any backfire effects even in the more vaccine-skeptical participants. Furthermore, the multi-item measures implemented in the present study to assess misinformation reliance likely provided a more reliable measure than the single-item measures utilized by Pluviano et al. It has been suggested that this lack of reliability may be the primary mechanism driving observed backfire effects, and in fact, to the best of our knowledge, *all* backfire effects reported with vaccine-related stimuli have been elicited using single-item measures [23,33]. Thus, Pluviano et al.'s finding may have simply been a false-positive, given that their sample size was significantly lower than the sample size in the present study (for a similar case, see [32]).

Formatted: Font: 12 pt

A clear applied implication from this research is therefore that the hesitancy surrounding repetition of misinformation during correction is largely unwarranted. In light of the broader literature, the repetition of misinformation within a correction may actually be beneficial rather than harmful. Repetition may increase the salience of the correction while also facilitating processes for conflict resolution and knowledge revision [13,27,63,71]. Although contemporary guidelines for debunking myths have already recognized this [8,9], the current study provides further evidence of the efficacy of corrections that repeat the to-be-corrected misinformation. Misinformation correction should therefore not be avoided because of fear of backfire effects, especially when it comes to important topics such as vaccinations—in the current pandemic, there is clear opportunity for our findings to be applied to misinformation regarding COVID-19 vaccinations. However, unnecessary repetition of misinformation should still be avoided as there is a risk that it will enhance familiarity without any added benefit [8,31,32]. Moreover, there will be situations in which misinformation should not be corrected at all, to avoid amplifying a disinformant and adopting their framing of an issue, or where the misinformation has little traction and thus presents low risk of harm [8,9,32,72].

Some limitations of the present research should be acknowledged. Our sample was an online sample that was relatively low in skepticism towards childhood vaccines. This was not a major concern for the present research because its main focus was on familiarity effects, which should occur independent of vaccine attitudes. However, future studies might consider using a sample from a more vaccine-skeptical population, as it is known that those with strongly-held beliefs can be motivated to defend them, potentially weakening the effectiveness of misinformation corrections in some circumstances (but see [49] for evidence to the contrary and further discussion). From an applied perspective, focusing on vaccine-skeptical individuals would be useful because it is this very population that needs to be

engaged if they are to be motivated to vaccinate. Another limitation is that only vaccination intention was measured, rather than actual uptake behaviour. As mentioned earlier, intentions do not consistently translate into action, and there are a range of factors beyond intention that determine and contribute to behaviour execution [66,67]. It is therefore recommended that future research investigate the uptake of healthcare behaviours following misinformation correction. Finally, we acknowledge some misalignment between intervention materials in some conditions and the measures obtained; for example, the fear appeal did not specifically relate to belief in the vaccine-autism link. As this was a direct replication, this was largely outside of our control. However, such misalignment might represent a threat to internal validity. For example, we might reasonably assume that all conditions were affected equally be the demand characteristics of this study; however, this may not actually be the case given the abovementioned misalignment in some conditions. Future research should therefore re-assess the interventions and their relative efficacy in more targeted studies.

Conclusion

This study sought to replicate the familiarity and fear-related backfire effects reported by Pluviano et al. (2017) [10]. We found no evidence to support the notion that misinformation repetition or fear appeals cause backfire effects. This suggests that the findings reported by Pluviano et al. were either worldview-driven or an artefact. This highlights the importance of reproducibility in psychological science [73]. The only intervention successful in reducing vaccine misconceptions was a myths-vs.-facts approach that repeated the to-be-corrected misinformation and juxtaposed it with alternative factual information. It is thus recommended that this approach is used to proactively counter vaccination misinformation where it is encountered.

Acknowledgments

We thank Charles Hanich for research assistance and Sara Pluviano for publicly making her materials available and kindly providing the child images that were no longer accessible online.

References

1. Lazer DMJ, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, et al. The science of fake news. Science. 2018;359:1094-6. https://doi.org/10.1126/science.aao2998
2. Tay LQ, Hurlstone MJ, Kurz T, Ecker UKH. A comparison of prebunking and debunking interventions for implied versus explicit misinformation. Brit J Psych. 2022;3:591-607. https://doi.org/10.1111/bjop.12551
3. Roozenbeek J, Schneider CR, Dryhurst S, Kerr J, Freeman ALJ, Recchia G, et al. Susceptibility to misinformation about COVID-19 around the world. R Soc Open Sci. 2020;7201199201199. https://doi.org/10.1098/rsos.201199
4. Poland GA, Spier R. Fear, misinformation, and innumerates: How the Wakefield paper, the press, and advocacy groups damaged the public health. Vaccine. 2010;28: 2361-2. https://doi.org/10.1016/j.vaccine.2010.02.052
5. Loomba S, de Figueiredo A, Piatek S, de Graaf K, Larson HJ. Measuring the impact of exposure to COVID-19 vaccine misinformation on vaccine intent in the UK and US. Nat Hum Behav. 2021;5:337-48. https://doi.org/10.1038/s41562-021-01056-1
6. Chan MPS, Jones CR, Jamieson KH, Albarracín D. Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. Psych Sci. 2017;28:1531-46. https://doi.org/10.1177/0956797617714579

7. Walter N, Tukachinsky R. 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? Comm Res. 2020;47:155-77. https://doi.org/10.1177/0093650219854600
8. Ecker UKH, Lewandowsky S, Cook J, Schmid P, Fazio LK, Brashier N, et al. The psychological drivers of misinformation belief and its resistance to correction. Nat Rev Psych. 2022;1:13-29. https://doi.org/10.1038/s44159-021-00006-y
9. Lewandowsky S, Cook J, Ecker UKH, Albarracín D, Amazeen MA, Kendeou P, et al. The Debunking Handbook 2020. https://sks.to/db2020
10. Pluviano S, Watt C, Della Sala SD. Misinformation lingers in memory: Failure of three pro-vaccination strategies. PLOS ONE. 2017;12:e0181640. https://doi.org/10.1371/journal.pone.0181640
11. Yonelinas AP. The nature of recollection and familiarity: A review of 30 years of research. J Mem Lang. 2002;46:441-517. https://doi.org/10.1006/jmla.2002.2864
12. Ecker UKH, Lewandowsky S, Tang DTW. Explicit warnings reduce but do not eliminate the continued influence of misinformation. Mem Cogn. 2010;38:1087-1100. https://doi.org/10.3758/MC.38.8.1087
13. Swire B, Ecker UKH, Lewandowsky S. The role of familiarity in correcting inaccurate information. J Exp Psych Learn Mem Cogn. 2017;43:1948-61. https://doi.org/10.1037/xlm0000422
14. Begg IM, Anas A, Farinacci S. Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. J Exp Psych Gen. 1992;121:446-58. https://doi.org/10.1037/0096-3445.121.4.446

15. Dechêne A, Stahl C, Hansen J, Wänke M. The truth about the truth: A meta-analytic review of the truth effect. Pers Soc Psych Rev. 2010;14:238-57. https://doi.org/10.1177/1088868309352251
16. Pennycook G, Cannon TD, Rand DG, Cowan N. Prior exposure increases perceived accuracy of fake news. J Exp Psych Gen. 2018;147:1865-80. https://doi.org/10.1037/xge0000465
17. Fazio LK, Brashier NM, Payne BK, Marsh EJ. Knowledge does not protect against illusory truth. J Exp Psych Gen. 2015;144:993-1002. https://doi.org/10.1037/xge0000098
18. Fazio LK. Repetition increases perceived truth even for known falsehoods. Collabra Psychol. 2020;6:38. doi: https://doi.org/10.1525/collabra.347
19. Lewandowsky S, Ecker UKH, Seifert CM, Schwarz N, Cook J. Misinformation and its correction: Continued influence and successful debiasing. Psych Sci Publ Int. 2012;13:106-31. https://doi.org/10.1177/1529100612451018
20. Schwarz N, Sanna LJ, Skurnik I, Yoon C. Metacognitive experiences and the intricacies of setting people straight: Implications for debiasing and public information campaigns. Adv Exp Soc Psych. 2007;39:127-61. https://doi.org/10.1016/S0065-2601(06)39003-X
21. Cook J, Lewandowsky S. The Debunking Handbook. 2011; University of Queensland. http://sks.to/debunk
22. Schwarz N, Newman E, Leach W. Making the truth stick & the myths fade: Lessons from cognitive psychology. Behav Sci Pol. 2016;2:85-95. https://doi.org/10.1353/bsp.2016.0009

23. Swire-Thompson B, DeGutis J, Lazer D. Searching for the backfire effect: Measurement and design considerations. J Appl Res Mem Cognit. 2020;9:286-99. https://doi.org/10.1016/j.jarmac.2020.06.006
24. Peter C, Koch T. When debunking scientific myths fails (and when it does not): The backfire effect in the context of journalistic coverage and immediate judgments as prevention strategy. Sci Comm. 2016;38:3-25. https://doi.org/10.1177/1075547015613523
25. Skurnik I, Yoon C, Park DC, Schwarz N. How warnings about false claims become recommendations. J Cons Res. 2005;31:713-24. https://doi.org/10.1086/426605
26. Cameron KA, Roloff ME, Friesema EM, Brown T, Jovanovic BD, Hauber S, et al. Patient knowledge and recall of health information following exposure to "facts and myths" message format variations. Pat Edu Couns. 2013;92,381-7. https://doi.org/10.1016/j.pec.2013.06.017
27. Ecker UKH, Lewandowsky S, Swire B, Chang D. Correcting false information in memory: Manipulating the strength of misinformation encoding and its retraction. Psychon Bull Rev. 2011;18:570-8. https://doi.org/10.3758/s13423-011-0065-1
28. Ecker UKH, Hogan JL, Lewandowsky S. Reminders and repetition of misinformation: Helping or hindering its retraction? J Appl Res Mem Cognit. 2017;6:185-92. https://doi.org/10.1016/j.jarmac.2017.01.014
29. Ecker UKH, O'Reilly Z, Reid JS, Chang EP. The effectiveness of short-format refutational fact-checks. Brit J Psych. 2020;111:36-54. https://doi.org/10.1111/bjop.12383
30. Bastin C, van der Linden M. The contribution of recollection and familiarity to recognition memory: A study of the effects of test format and aging. Neuropsych. 2003;17:14-24. https://doi.org/10.1037//0894-4105.17.1.14

31. Autry, KS, Duarte, SE. Correcting the unknown: Negated corrections may increase belief in misinformation. Appl Cognit Psychol. 2021;35:960-75. https://doi.org/10.1002/acp.3823
32. Ecker UKH, Lewandowsky S, Chadwick M. Can corrections spread misinformation to new audiences? Testing for the elusive familiarity backfire effect. Cogn Res Princ Implic. 2020;5:41. https://doi.org/10.1186/s41235-020-00241-6
33. Swire-Thompson B, Miklaucic N, Wihbey JP, Lazer D, DeGutis J. Backfire effects after correcting misinformation are strongly associated with reliability. J Exp Psych Gen. 2022;151:1655-65. https://doi.org/10.31234/osf.io/e3pvx
34. Pluviano S, Watt C, Ragazzini G, Della Sala SD. Parents' beliefs in misinformation about vaccines are strengthened by pro-vaccine campaigns. Cognit Proc. 2019;20: 325-31. https://doi.org/10.1007/s10339-019-00919-w
35. Johnson HM, Seifert CM. Sources of the continued influence effect: When misinformation in memory affects later inferences. J Exp Psych Learn Mem Cognit. 1994;20:1420-36. https://doi.org/10.1037/0278-7393.20.6.1420
36. Lipkus IM, Hollands JG. The visual communication of risk. J Nat Canc Inst Monogr. 1999;25:149-63. https://doi.org/10.1093/oxfordjournals.jncimonographs.a024191
37. Dixon GN, McKeever BW, Holton AE, Clarke C, Eosco G. The power of a picture: Overcoming scientific misinformation by communicating weight-of-evidence information with visual exemplars. J Comm. 2015;65:639-59. https://doi.org/10.1111/jcom.12159
38. Danielson RW, Sinatra GM, Kendeou P. Augmenting the refutation text effect with analogies and graphics. Disc Proc. 2016;53:392-414. https://doi.org/10.1080/0163853X.2016.1166334

39. Nyhan B, Reifler J. The roles of information deficits and identity threat in the prevalence of misperceptions. J Elect Pub Opin Part. 2019;29:222-44. https://doi.org/10.1080/17457289.2018.1465061

40. van der Linden SL, Leiserowitz AA, Feinberg GD, Maibach EW. How to communicate the scientific consensus on climate change: Plain facts, pie charts or metaphors? Climatic Change. 2014;126:255-62. https://doi.org/10.1007/s10584-014-1190-4

41. Witte K, Allen M. A meta-analysis of fear appeals: Implications for effective public health campaigns. Health Edu Behav. 2000;27:591-615. https://doi.org/10.1177/109019810002700506

42. Tannenbaum MB, Hepler J, Zimmerman RS, Saul L, Jacobs S, Wilson K, Albarracín D. Appealing to fear: A meta-analysis of fear appeal effectiveness and theories. Psych Bull. 2015;141:1178-1204. https://doi.org/10.1037/a0039729

43. MacFarlane D, Hurlstone MJ, Ecker UKH. Countering demand for unsupported health remedies: Do consumers respond to risks, lack of benefits, or both? Psych Health. 2020;12:593-611. https://doi.org/10.1080/08870446.2020.1774056

44. Joffe H. The power of visual material: Persuasion, emotion and identification. Diogenes. 2008;55:84-93. https://doi.org/10.1177/0392192107087919

45. Nyhan B, Reifler J, Sean R, Freed GL. Effective messages in vaccine promotion: a randomized trial. Pediatrics. 2014;133:e835-42. https://doi.org/10.1542/peds.2013-2365

46. Goldenberg J. Misattribution of arousal. In: Baumeister RF, Vohs KD, editors. Encyclopedia of Social Psychology. Vol. 2. Gale eBooks; 2007. pp. 581-583. https://go.gale.com/ps/i.do?p=GVRL&u=uwa&id=GALE%7CCX2661100341&v=2.1&it=r

47. Nyhan B, Reifler J. When corrections fail: The persistence of political misperceptions. Polit Behav. 2010;32:303-30. https://doi.org/10.1007/s11109-010-9112-2

48. Ecker UKH, Ang LC. Political attitudes and the processing of misinformation corrections. Polit Psych. 2019;40:241-60. https://doi.org/10.1111/pops.12494

49. Ecker UKH, Sze BKN, Andreotta M. Corrections of political misinformation: no evidence for an effect of partisan worldview in a US convenience sample. Phil Trans Roy Soc B. 2021;376:20200145. https://doi.org/10.1098/rstb.2020.0145

50. Wood T, Porter E. The elusive backfire effect: Mass attitudes' steadfast factual adherence. Polit Behav. 2019;41:135-63. https://doi.org/10.1007/s11109-018-9443-y

51. Haglin K. The limitations of the backfire effect. Res Politics. 2017;4:1-5. https://doi.org/10.1177/2053168017716547

52. Schmid P, Betsch C. Effective strategies for rebutting science denialism in public discussions. Nat Hum Behav. 2019;3:931–9. https://doi.org/10.1038/s41562-019-0632-4

53. Nyhan B, Reifler J. Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information. Vaccine. 2015;33:459-64. https://doi.org/10.1016/j.vaccine.2014.11.017

54. Horne Z, Powell D, Hummel JE, Holyoak KJ. Countering antivaccination attitudes. Proc Natl Acad Sci U.S.A. 2015;112:10321-4. https://doi.org/10.1073/pnas.1504019112

55. de Figueiredo A, Simas C, Karafillakis E, Paterson P, Larson HJ. Mapping global trends in vaccine confidence and investigating barriers to vaccine uptake: a large-scale retrospective temporal modelling study. Lancet. 2020;396:898-908. https://doi.org/10.1016/S0140-6736(20)31558-0

56. Giambi C, Fabiani M, D'Ancona F, et al. Parental vaccine hesitancy in Italy – Results from a national survey. Vaccine. 2018;36:779-87. https://doi.org/10.1016/j.vaccine.2017.12.074

57. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behav Res Meth. 2009;41:1149-60. https://doi.org/10.3758/BRM.41.4.1149

58. Martin LR, Petrie KJ. Understanding the dimensions of anti-vaccination attitudes: the Vaccination Attitudes Examination (VAX) scale. Annals Behav Med. 2017;51:652-60. https://doi.org/10.1007/s12160-017-9888-y

59. Wood L, Smith M, Miller CB, O'Carroll RE. The internal consistency and validity of the Vaccination Attitudes Examination scale: A replication study. Annals Behav Med. 2019;53:109-14. https://doi.org/10.1093/abm/kay043

60. Hoaglin DC, Iglewicz B. Fine-tuning some resistant rules for outlier labeling. J Amer Stat Assoc. 1987;82:1147-9. https://doi.org/10.1080/01621459.1987.10478551

61. Rubin M. When to adjust alpha during multiple testing: A consideration of disjunction, conjunction, and individual testing. Synthese. 2021;199:10969–1000. https://doi.org/10.1007/s11229-021-03276-4

62. Turk-Browne NB, Jungé JA, Scholl BJ. The automaticity of visual statistical learning. J Exp Psychol Gen. 2005;134:552-64. https://doi.org/10.1037/0096-3445.134.4.552

63. Swire-Thompson B, Cook J, Butler LH, Sanderson JA, Lewandowsky S, Ecker UKH. Correction format has a limited role when debunking misinformation. Cogn Res Princ Implic. 2021;6:83. https://doi.org/10.1186/s41235-021-00346-6

64. Winters M, Oppenheim B, Sengeh P, et al. Debunking highly prevalent health misinformation using audio dramas delivered by WhatsApp: Evidence from a

randomised controlled trial in Sierra Leone. BMJ Glob Health. 2021;6:e006954. https://doi.org/10.1136/bmjhg-2021-006954

65. Martel C, Mosleh M, Rand D. You're definitely wrong, maybe: Correction style has minimal effect on corrections of misinformation online. Media Comm. 2021;9:120-33. https://doi.org/10.17645/mac.v9i1.3519
66. McEachan R, Conner M, Taylor N, Lawton R. Prospective prediction of health-related behaviours with the theory of planned behaviour: A meta-analysis. Health Psychol Rev. 2011;5:97-144. https://doi.org/10.1080/17437199.2010.521684
67. Sheeran P, Webb TL. The intention-behavior gap. Soc Personal Psychol Compass. 2016;10:503-18. https://doi.org/10.1111/spc3.12265
68. Pluviano S, Della Sala S, Watt C. The effects of source expertise and trustworthiness on recollection: The case of vaccine misinformation. Cogn Process. 2020;21:321-30. https://doi.org/10.1007/s10339-020-00974-8
69. Swire-Thompson B, Ecker UKH, Lewandowsky S, Berinsky A. They might be a liar but they're my liar: Source evaluation and the prevalence of misinformation. Polit Psychol. 2020;41:21-34. https://doi.org/10.1111/pops.12586
70. Rezza G. Mandatory vaccination for infants and children: the Italian experience, Pathog Glob Health. 2019;113:291-6. https://doi.org/10.1080/20477724.2019.1705021
71. Kendeou P, Walsh EK, Smith ER, O'Brien EJ. Knowledge revision processes in refutation texts. Discourse Process. 2014;51:374-97. https://doi.org/10.1080/0163853X.2014.913961
72. Looi MK. Sixty seconds on…the NyQuil chicken challenge. BMJ. 2022;378:o2298. http://dx.doi.org/10.1136/bmj.o2298

73. Redish AD, Kummerfeld E, Morris RL, Love AC. Opinion: Reproducibility failures are essential to scientific inquiry. Proc. Natl. Acad. Sci. U.S.A. 2018;115:5042-6. https://doi.org/10.1073/pnas.1806370115