PAPER IN PRESS AT THE JOURNAL OF APPLIED RESEARCH IN MEMORY AND COGNITION

Social Endorsement Influences the Continued Belief in Corrected Misinformation

Lucy H. Butler, Nicolas Fay & Ullrich K. H. Ecker

School of Psychological Science, University of Western Australia

Word Count: 6,645 (Main text incl. figure captions and tables); 2,940 (Introduction and Discussion)

Acknowledgements: This research was supported by an Australian Government Research Training Program Scholarship to the first author, an Office of National Intelligence and Australian Research Council grant NI210100224 to the second author, and an Australian Research Council grant FT190100708 to the last author.

Address correspondence to: Lucy Butler, School of Psychological Science (M304), University of Western Australia, 35 Stirling Hwy, Perth 6009, Australia. Telephone: +618 6488 7850; email: lucy.butler@research.uwa.edu.au.

Abstract

Reliance on misinformation often persists in the face of corrections. However, the role of social factors on people's reliance on corrected misinformation has received little attention. In two experiments, we investigated the extent to which social endorsement of misinformation and corrections influences belief updating. In both experiments misinformation and fact-checks were presented as social-media posts, and social endorsement was manipulated via the number of "likes". In Experiment 1, social endorsement of the initial misinformation had a significant influence on belief; participants believed misinformation with high social endorsement more than misinformation with low endorsement. This effect was observed pre-fact-check and post-fact-check. High social endorsement of the fact-checks was associated with reduced misinformation belief; however, evidence for the persistence of this effect was mixed. These findings were replicated in Experiment 2. Our findings indicate that social endorsement can moderate our beliefs in misinformation and the fact-checks designed to correct these beliefs.

Keywords: Misinformation; Continued influence effect; Social endorsement; Belief updating

General Audience Summary

Misinformation can be created and spread on social-media platforms with relative ease. Additionally, unlike traditional media, common cues of information credibility, such as source expertise and trustworthiness, are often unavailable on social-media platforms. As such, determining what to believe versus what to disregard can be difficult. To reduce the effort required to assess information credibility, people may rely on mental shortcuts (or heuristics). Specifically, it is well established that people often look to others, or the majority, when making decisions about what to believe and how to behave. As such, people may rely on engagement metrics on social media (e.g., the number of "likes" or "shares" a post has) to gauge information credibility. Across two experiments we investigated how level of social endorsement (specifically, whether a social-media post has a high or low number of "likes") influenced the extent to which people believed misinformation, and updated their belief based on subsequent fact-checks. We found that people had greater belief in misinformation with a high versus low level of likes, even after they received a fact-check. Further, fact-checks with a high number of likes reduced belief in misinformation more than fact-checks with a low number of likes. However, evidence for the persistence of the effect of fact-check endorsement on belief updating was mixed. These findings suggest that people may rely on engagement metrics, specifically number of likes, when appraising misinformation on social media. This influence of endorsement information may be reason for concern, particularly given that engagement information can be maliciously manipulated, and misleading and conspiratorial claims often have characteristics designed to enhance their level of endorsement.

Social Endorsement Influences the Continued Belief in Corrected Misinformation

Misinformation—inaccurate or misleading information that is presented as factual—permeates all aspects of the modern media environment (Vosoughi et al., 2018). Due to the negative influence misinformation can have on individuals and society (e.g., vaccine misinformation; Donzelli et al., 2018), considerable effort has gone into strategies to correct misinformation (e.g., Lewandowsky et al., 2020). However, the ability of corrections to reduce false beliefs is limited: Although corrections can reduce belief in misinformation, they do not eliminate its influence; people often continue to rely on corrected misinformation in their reasoning. This phenomenon is known as the continued influence effect (CIE; Ecker, Lewandowsky et al., 2022; Johnson & Seifert, 1994). To improve the impact of corrective efforts, it is necessary to understand the factors underlying the CIE.

Prior research has focused on the cognitive basis of the CIE, showing that continued reliance on misinformation is due to failures in memory updating and retrieval (e.g., Ecker et al., 2017; Gordon et al., 2019; Jia et al., 2020; Kendeou et al., 2014; Swire et al., 2017). While social factors are also likely to be important to the CIE (Ecker et al., 2022; Smith & Semin, 2007; Van Bavel et al., 2021), the role they play has received little attention. Here we directly assess how social endorsement influences misinformation belief and the CIE.

One social factor known to influence misinformation susceptibility is source credibility. A recent meta-analysis found that credibility of a misinformation source can contribute to false beliefs (Walter & Tukachinsky, 2020). Although the meta-analysis found no significant effect of correction-source credibility on belief updating, there is some evidence that correction-source trustworthiness influences correction acceptance (Ecker & Antonio, 2021; Guillory & Geraci, 2013). However, with the sheer volume of user-generated content alongside the source ambiguity on social media, information credibility can be difficult to gauge (Flanagin, 2017). Consequently, people may use heuristic cues, such as

normative information derived from social-endorsement levels (i.e., the bandwagon heuristic; Sundar, 2008) to assess information credibility (Hornsey et al., 2003; Metzger et al., 2010).

Normative information, specifically majority behaviours or beliefs, can influence individual decision making (Asch, 1956; Kendal et al., 2018). Majority influence can occur due to the assumption that others have access to information the individual does not, or because individuals are motivated to conform to avoid social exclusion (Kaplan & Miller, 1987). Reliance on social norms arises when there is a high degree of uncertainty surrounding information quality, such as when there is conflicting evidence (Cialdini & Goldstein, 2004), when the credibility of the information source is unknown (Mena et al., 2020), or when individuals do not have the capacity to critically appraise the information available (Metzger & Flanagin, 2013). In such situations, if misinformation is perceived as normative and corrective information as non-normative, correction effectiveness may be impeded.

Social endorsement metrics are increasingly present in digital media (Kim, 2021; Metzger et al., 2010). Specifically, engagement metrics that signal endorsement, in the form of "likes" or "shares", are ubiquitous on social-media (Peters et al., 2013). Given the high information load and relative absence of source-credibility cues on these sites (Qiu et al., 2017), it is plausible that people may rely on these endorsement cues to assess information veracity (Mena et al., 2020; Metzger et al., 2010). Prior research has provided some evidence for the role of social endorsement on misinformation acceptance, finding that people are more susceptible to low-quality information when it is associated with a high (compared to low) level of social endorsement (i.e., number of "likes"; Avram et al., 2020). Recently, Vlasceanu and Coman (2021) found a comparable effect of social endorsement on *correction* effectiveness, with high (vs. low) endorsement of corrections associated with greater belief updating. This suggests that people may be similarly influenced by the social endorsement of misinformation and corrections, and that if a piece of misinformation receives a greater level

of endorsement than a correction, as is typical with viral misinformation (Bode & Vraga, 2018), a particularly strong CIE may arise.

Prior research has not examined misinformation and correction endorsement within the same study, and as such their relative influence, and any potential interaction between the two factors, is unknown. Additionally, Avram and colleagues (2020) used behavioural measures (i.e., willingness to "share" or "like" the misinformation), which provide insight into people's tendency to propagate misinformation, but does not provide insight into misinformation beliefs (see Van Bavel et al., 2021). Across two experiments, we assessed the influence of social endorsement on misinformation and correction acceptance using a direct assessment of participants' belief in misinformation. By using a fully-crossed design we could assess: (1) if the effect of misinformation endorsement on belief is sustained following a correction, and (2) whether misinformation and correction endorsements interact. It was hypothesised that belief in misinformation would be greater—both before and after a correction—when it is associated with a high (vs. low) level of social endorsement. It was also hypothesised that corrections associated with a high (vs. low) level of endorsement would yield more belief updating (i.e., a reduced CIE).

Experiment 1

Method

Experiment 1 presented participants with false claims (as well as true filler claims) in the format of social-media posts. Each claim was later fact-checked in a separate post. All claims and fact-checks were associated with either a high or low level of social endorsement (i.e., number of "likes"). Belief in the claims was measured at three time-points: immediately following claim exposure (belief rating 1), immediately following exposure to the associated fact-check (belief rating 2), and at the end of the experiment, following a distraction task (belief rating 3). Thus, the experiment used a 2 × 2 × 3 within-subjects design, with factors

claim endorsement (high, low), correction endorsement (high, low), and time (1: pre-fact-check; 2: post-fact-check; 3: post-delay). In addition, claim-congruent inferential reasoning was measured at time-point 3 using inference questions that indirectly assessed reliance on misinformation. Note that the inclusion of true claims means that technically the design included an additional within-subjects factor of claim veracity (false, true); however, the focus of the research is false claims, and true claims (and associated affirmations) were only included to ensure some balance in the participant experience and to avoid strong demand characteristics (i.e., the perception that all claims are false and all fact-checks are meant to reduce beliefs). Analyses regarding true filler claims were conducted separately and for exploratory purposes only. Results for true filler claims are briefly described in the Results section; however, full analyses are provided in Supplement B, available at https://osf.io/vt7p6/.

Participants

U.S.-based adult participants were recruited through the crowdsourcing platform Prolific. An a-priori power analysis (using G*Power; Faul et al., 2007) suggested a minimum sample size of 352 to detect a small effect of $f = 0.15$ at $\alpha = 0.05$ and $1 - \beta = 0.8$. To ensure adequate power following exclusions, a total of 382 participants were recruited. Participants were excluded based on the following a-priori criteria: self-reported English proficiency rated as only “fair” or “poor” ($n = 1$); self-reported lack of effort ($n = 0$); completion time < 10 min ($n = 0$); uniform responding (identical response to ≥ 90 out of 120 items; $n = 0$); inconsistent responding (where the difference between mean responses to standard vs. reverse-coded inference questions was identified as an outlier, using the outlier-labelling rule with a 1.5 multiplier; $n = 15$). This resulted in a final sample size of $N = 366$; the sample comprised 99 males, 255 females, and 12 non-binary individuals; mean age was $M = 28.97$ years ($SD = 9.94$, age range: 18-72).

Materials

To minimise the influence of worldview effects on misinformation or correction processing (Walter & Tukachinsky, 2020), politically neutral claims were used (e.g., the false claim that *"Students learn best when teaching styles are matched to their learning style"*). A total of 16 false claims and 8 true filler claims were selected from an initial pool of 42 false claims and 20 true claims rated by a separate sample of $N = 100$ participants. False claims were selected that had a relatively high level of believability (whereas true filler claims were selected that had somewhat lower believability), to avoid floor (ceiling) effects during belief updating, whilst maintaining a moderate and comparable level of claim familiarity across claims. To ensure that high levels of endorsement were plausible, claims were also rated on the degree to which a high level of social-media engagement would be perceived as surprising and thus unrealistic; claims were excluded if they were rated > 5 (on a 0-10 scale) on this dimension. See Supplement A for full selection details.

Claims were associated with one of four fictitious source accounts. Claim-source pairings were fixed; they maintained comparable average claim believability (based on the pilot data) across sources. Corresponding fact-checks were developed for each of the 16 false claims (and 8 true filler claims) and displayed in an analogous social-media format (see Figure 1 for an example claim and fact-check). Fact-checks were associated with four different fictitious sources, and pairings of fact-checks and sources were fixed. Fact-check style varied slightly between sources; however, all fact-checks repeated the initial misinformation with an accompanying false tag (see Supplement A for details). Each of the 16 possible combinations of false-claim and fact-check sources was presented once to each participant.

Figure 1

Example False Claim and Associated Fact-Check

Note. The false claim (left) was presented before belief rating 1; the associated fact-check (right) was presented before belief rating 2. The example illustrates the low/high condition (low level of claim endorsement; high level of fact-check endorsement).

Each claim and each fact-check was associated with a high or low number of "likes". For each post, the number of likes was sampled from a normal distribution, with $M = 10$ ($SD = 4$) for low endorsement, and $M = 1{,}000$ ($SD = 200$) for high endorsement. The range was determined arbitrarily but designed to maximise differences whilst maintaining plausibility. Endorsement levels of claims and fact-checks were fully crossed, resulting in four condition combinations (high/high; high/low; low/high; low/low). Assignment of claims to conditions was fully counterbalanced across participants.

Measures. Claim belief was measured through direct ratings on a Likert scale ranging from 0 (“certainly false”) to 10 (“certainly true”). False-claim-specific reasoning was measured through 32 inferential-reasoning questions (16 additional inferential-reasoning questions relating to the true filler claims were also included), with two questions (one reverse-coded) per claim. Each question presented participants with a statement related to the claim (an example reverse-coded item for the false claim shown in Figure 1 is *“Teachers should disregard learning styles when developing their curriculum”*) and participants rated their agreement on a Likert scale ranging from 0 (“strongly disagree”) to 10 (“strongly agree”). See Supplement A for all questions.

Procedure

The experiment was run using Qualtrics software (Qualtrics, Provo, UT). Participants initially received an ethics-approved information sheet, provided informed consent, and completed a short demographics questionnaire. All claims (including the eight true filler claims) were presented individually and in a randomised order. For each claim, participants rated their initial belief (pre-fact-check; belief rating 1) before being presented with the associated fact-check (a correction for the false claims; an affirmation for the true filler claims) and then providing a second belief rating, on a separate page (post-fact-check; belief rating 2), before moving to the next claim. This procedure was chosen to minimise interference and cognitive load placed on participants in an effort to isolate the specific effect of social endorsement. After the presentation of all the claims and fact-checks, participants completed a 1-minute distraction task (a word puzzle). This was followed by the inference questions (with question pairs presented in a randomised order) and the final belief rating of all claims (post-delay; belief rating 3). Claims at the final belief rating were displayed with no endorsement information. After completion of the final belief rating, participants were asked to self-report whether their data should be discarded due to lack of effort, before being

debriefed. The debrief informed participants that the corrections and affirmations were accurate to the best of our knowledge (and restated specifically which claims were inaccurate), whereas the endorsement information (i.e., the associated likes) was simulated. Median completion time was 18.6 minutes; participants were compensated £2.50 (approximately US$3.40).

Results

All data and additional analyses are available at https://osf.io/vt7p6/. Given our outcome variables were measured on a Likert scale, analyses were conducted using cumulative link (ordinal) mixed effects modelling (CLMM; McElreath, 2020), using the *clmm* function of the *ordinal* R package (Christensen, 2018).[1] Unlike ANOVAs, cumulative link mixed-modelling does not assume that there are proportionate intervals between scale options, or that data come from a continuous scale.

Prior to statistical analyses, the fixed effects of claim endorsement and fact-check endorsement were centred, and time (pre-fact-check, post-fact-check, post-delay) was factor-coded. Dependent variables of belief and inference score were coded as ordinal factors with 11 levels (0-10). For each analysis, the maximum random-effects structure justified by the experimental design was included where possible. The random effects included by-participant and by-stimuli (i.e., claims) random intercepts and by-participant and by-stimuli random slopes for claim endorsement (high, low), fact-check endorsement (high, low), and their interaction. Specification of the random effects structures are provided in the Supplement C.

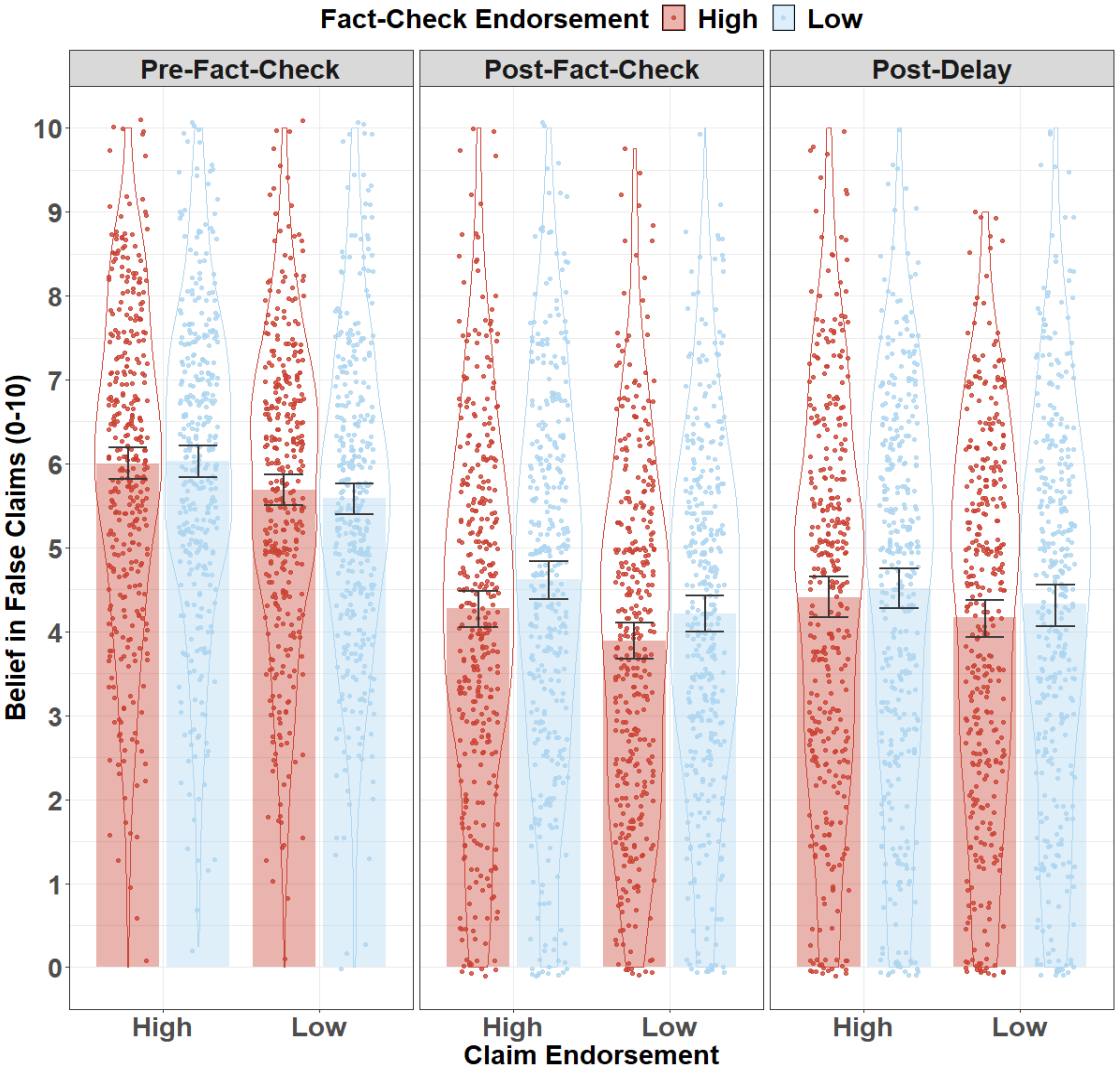
False Claims

[1] Our initial analysis plan specified ANOVAs for the primary analysis and CLMM as a secondary analysis. We report CLMM results in the main text based on reviewer feedback. ANOVA results were comparable and are provided in the Supplement B.

Belief Ratings. Mean false-claim belief ratings across time-points and conditions are presented in Figure 2. We first tested the effects of claim endorsement and fact-check endorsement at time 1 (pre-fact-check), to assess the immediate effect of claim endorsements, and to ensure fact-check endorsement conditions did not significantly differ prior to fact-check exposure. There was a fixed effect of claim endorsement, $\beta = .29$, $SE = .06$, $z = 5.26$, $p < .001$. As expected, there was no statistical evidence for an effect of fact-check endorsement, $\beta = .02$, $SE = .05$, $z = .32$, $p = .753$, or an interaction between claim endorsement and fact-check endorsement, $\beta = -.13$, $SE = .10$, $z = -1.38$, $p = .168$. The effect of claim endorsement indicates that belief in misinformation with high endorsement was greater than belief in misinformation with low endorsement.

Figure 2

Mean False-Claim Belief Ratings Across Time-points and Conditions in Experiment 1



Note. Error bars represent 95% confidence intervals.

Next, we examined the effect of claim endorsement and fact-check endorsement, and their potential interaction, from time-point 1 to 2. We were interested in the effect of the fact check on belief in the original claim, and the extent to which this was moderated by social endorsement information. There was a fixed effect of claim endorsement, $\beta = .30$, $SE = .07$, $z = 4.34$, $p < .001$, and time, $\beta = -1.13$, $SE = .03$, $z = -32.90$, $p < .001$, as well as a time × fact-check-endorsement interaction, $\beta = -.26$, $SE = .07$, $z = -4.02$, $p < .001$. The effect of claim endorsement mirrors that at time 1, and the effect of time indicates that belief in

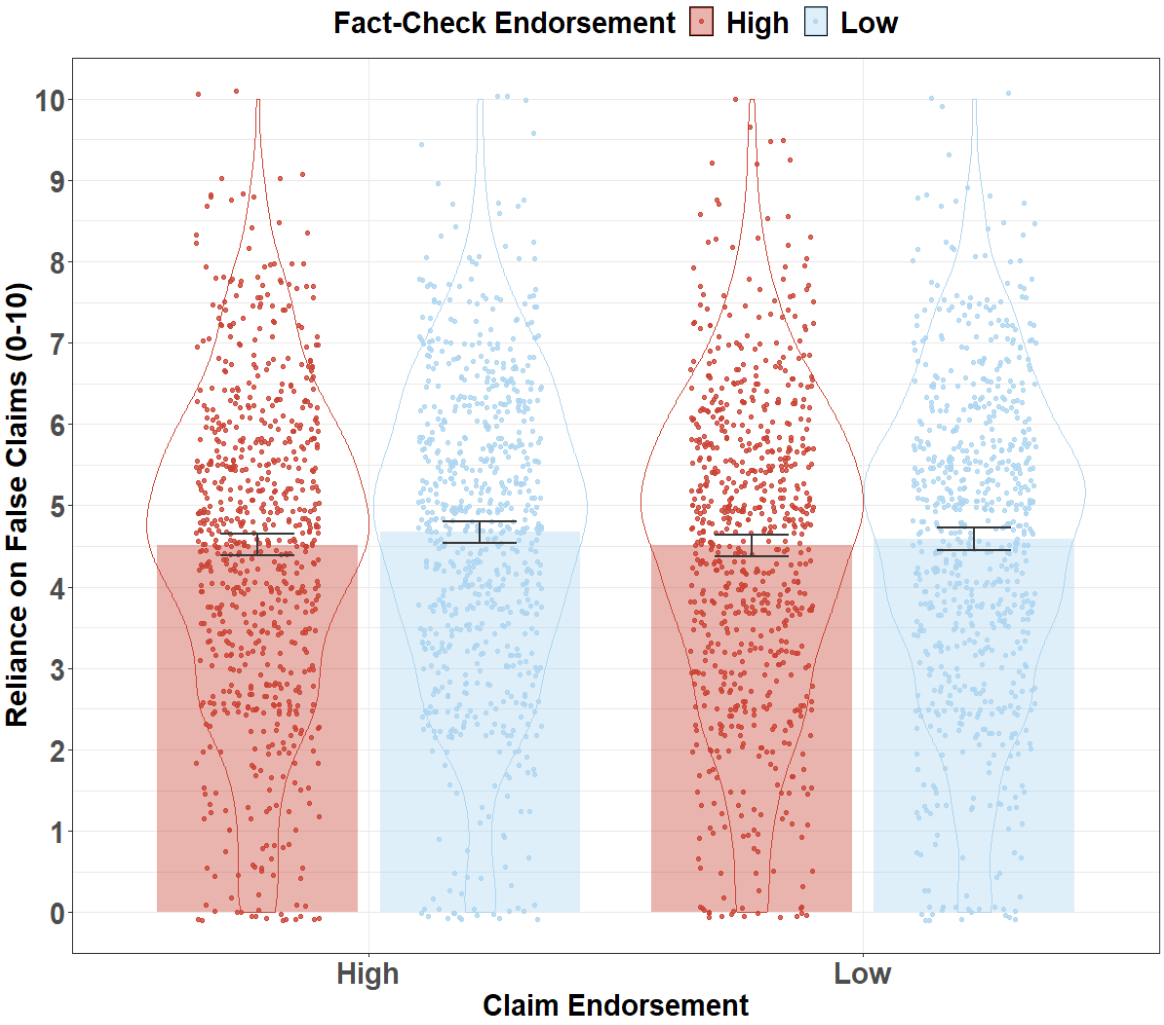
misinformation reduced following the fact-check. The interaction indicates that the effect of fact-check endorsement was significantly greater post-fact-check compared to pre-fact-check. To quantify this significant increase in the effect of fact-check endorsement, we assessed results of fact-check endorsement only at time-point 2 (post-fact-check). In line with predictions, there was a fixed effect of fact-check endorsement, $\beta = -.24$, $SE = .06$, $z = -4.25$, $p < .001$, with higher fact-check endorsement being associated with greater belief updating (i.e., lower belief in the false claims).

To assess whether the effects of claim and fact-check endorsement were maintained over time, claim belief was assessed across time-points 2 and 3. There was a fixed effect of claim endorsement, $\beta = .29$, $SE = .06$, $z = 5.23$, $p < .001$, and a fixed effect of fact-check endorsement, $\beta = -.24$, $SE = .05$, $z = -4.54$, $p < .001$, as well as time × claim-endorsement, $\beta = -.14$, $SE = .07$, $z = -2.17$, $p = .030$, and time × fact-check-endorsement, $\beta = .15$, $SE = .07$, $z = 2.30$, $p = .022$, interactions. To quantify the interaction effects, false claim belief was assessed at time 3. At time 3, there was a fixed effect of claim endorsement, $\beta = .15$, $SE = .05$, $z = 3.16$, $p = .002$, but no effect of fact-check endorsement, $\beta = -.09$, $SE = .05$, $z = -1.83$, $p = .067$, or interaction, $\beta = .03$, $SE = .10$, $z = .32$, $p = .754$. Thus, the claim-endorsement effect seen at time 2 was maintained at time 3, whereas the fact-check-endorsement effect was not maintained.

Inference Scores. Mean inference scores across conditions are presented in Figure 3. There was no statistical evidence of an effect of claim endorsement, $\beta = .03$, $SE = .04$, $z = .92$, $p = .359$. There was a significant fixed effect of fact-check endorsement, $\beta = -.09$, $SE = .05$, $z = -2.09$, $p = .036$, indicating that reliance on misinformation was lower (i.e., greater belief updating) when fact-checks had a high compared to low level of endorsement. There was no statistical evidence of a claim-endorsement × fact-check-endorsement interaction, $\beta = -.04$, $SE = .11$, $z = -.56$, $p = .737$.

Figure 3

Mean False-Claim Inference Scores Across Conditions in Experiment 1



Note. Error bars represent 95% confidence intervals.

True Claims

Analyses of true claims were exploratory and results are briefly described below; see Supplement B for full details. Belief in true claims was analysed in an identical fashion to belief in false claims. At all three time points, there was a fixed effect of claim endorsement, with higher endorsement associated with greater belief ratings (in line with the false-claims results). There was no claim-endorsement × time interaction effect on true-claim beliefs. There was a significant fact-check-endorsement × time interaction between times 1 and 2,

with greater fact-check endorsement associated with greater belief updating. There was also a significant fact-check-endorsement × time interaction between times 2 and 3: The effect of fact-check endorsement reduced between time-points 2 (post-fact-check) and 3 (post-delay); however, the effect of fact-check endorsement remained significant even at time 3. True-claim inference scores were also analysed in an identical manner to false-claim inference scores. There was a fixed effect of claim endorsement, indicating reliance on true information was higher when claims had a high compared to low level of endorsement.

Discussion

Experiment 1 found a clear effect of false-claim endorsement on misinformation beliefs; participants' beliefs were measurably higher when the misinformation had a high compared to low level of social endorsement, an effect observed prior to and following the fact-check. This finding is in line with our first hypothesis, and supports the notion that social endorsement influences people's susceptibility to misinformation, particularly in situations where other information credibility cues are absent.

In line with our second hypothesis, there was also a significant effect of fact-check endorsement on belief in corrected misinformation. Specifically, at time 2 belief updating was significantly greater when the fact-check was associated with a high compared to low level of social endorsement. Although higher fact-check endorsement was associated with greater belief updating at time 2, this effect reduced after a short delay, with no statistical evidence of an effect of fact-check endorsement on misinformation belief at time-point 3. Given the length of the delay (approx. 10-15 minutes on average), this suggests that level of fact-check endorsement may not have an enduring effect on belief updating. With this in mind, however, we note that the effect of social endorsement on indirect reliance on misinformation (i.e., inference scores) seemingly contradicts the effect on belief scores, with only fact-check endorsement yielding a significant result. On the one hand, the effect was

weak, and in the context of the high cognitive load it may reflect noise. On the other hand, it could represent a small but real effect, which could not be substantiated statistically in the analysis of belief ratings at time-point 3. This considered, the observed effect did appear to be driven by higher inference scores in the low-fact-check-endorsement condition, which is in line with predictions and the numerically higher belief ratings observed in this condition.

Experiment 2

The aim of Experiment 2 was twofold. First, it aimed to replicate the effects of claim endorsement and fact-check endorsement observed in Experiment 1. Second, by including a no-endorsement control condition, it aimed to determine if the endorsement effects observed in Experiment 1 were driven by high claim (fact-check) endorsements increasing belief (belief updating), low claim (fact-check) endorsements decreasing belief (belief updating), or a combination of the two.

The design and analysis plan were pre-registered (https://osf.io/wp3vt/registrations).[2] The core hypotheses remained identical to Experiment 1. Additional hypotheses were formed regarding the effects of endorsement compared to the no-endorsement control. Specifically, it was hypothesised that compared to the no-endorsement control condition, false-claim belief would be greater in the high-endorsement and lower in the low-endorsement condition. Further, it was hypothesised that compared to the no-endorsement control condition, belief updating would be greater (i.e., a smaller CIE) after fact-checks with high endorsement, and lower after fact-checks with low endorsement.

Method

[2] The pre-registered analysis plan specified ANOVAs for the primary analysis and CLMM as a secondary analysis. We diverge from this analysis plan, reporting CLMM results in the main text and ANOVA results, which were largely comparable, in the Supplement. Outcome discrepancies are noted in the main text.

Participants

Sampling for the experimental conditions was based on the power analysis of Experiment 1. To ensure approximately equal sample sizes across the control and experimental conditions, the minimum sample size for the study was set to 704. A total of 733 adult U.S. participants were sampled to ensure adequate sample size post exclusions. No participants were excluded due to low English proficiency, lack of effort, completion time < 10 minutes, or uniform responding. Twenty-one participants were excluded due to inconsistent responding (in line with the criterion used in Experiment 1), resulting in a total sample size of $N = 712$ ($n = 355$ in the experimental group; $n = 357$ in the control group). The sample consisted of 442 females, 257 males, and 13 non-binary individuals; mean age was $M = 40.18$ years ($SD = 13.91$, age range: 18-84).

Materials and Procedure

Materials and procedure were identical to Experiment 1. The only exception was that the claims in the control condition did not contain any endorsement information (i.e., no like symbol or like count), and there was no mention of endorsement information in the instructions in this condition.

Results

Data were prepared and results for the experimental conditions were analysed in line with Experiment 1. Data and supplementary analyses are available at https://osf.io/vt7p6/.

False Claims

Belief Ratings. Mean false-claim belief ratings across time-points and experimental conditions are presented in Figure 4. The same statistical models used in Experiment 1 were used in Experiment 2.

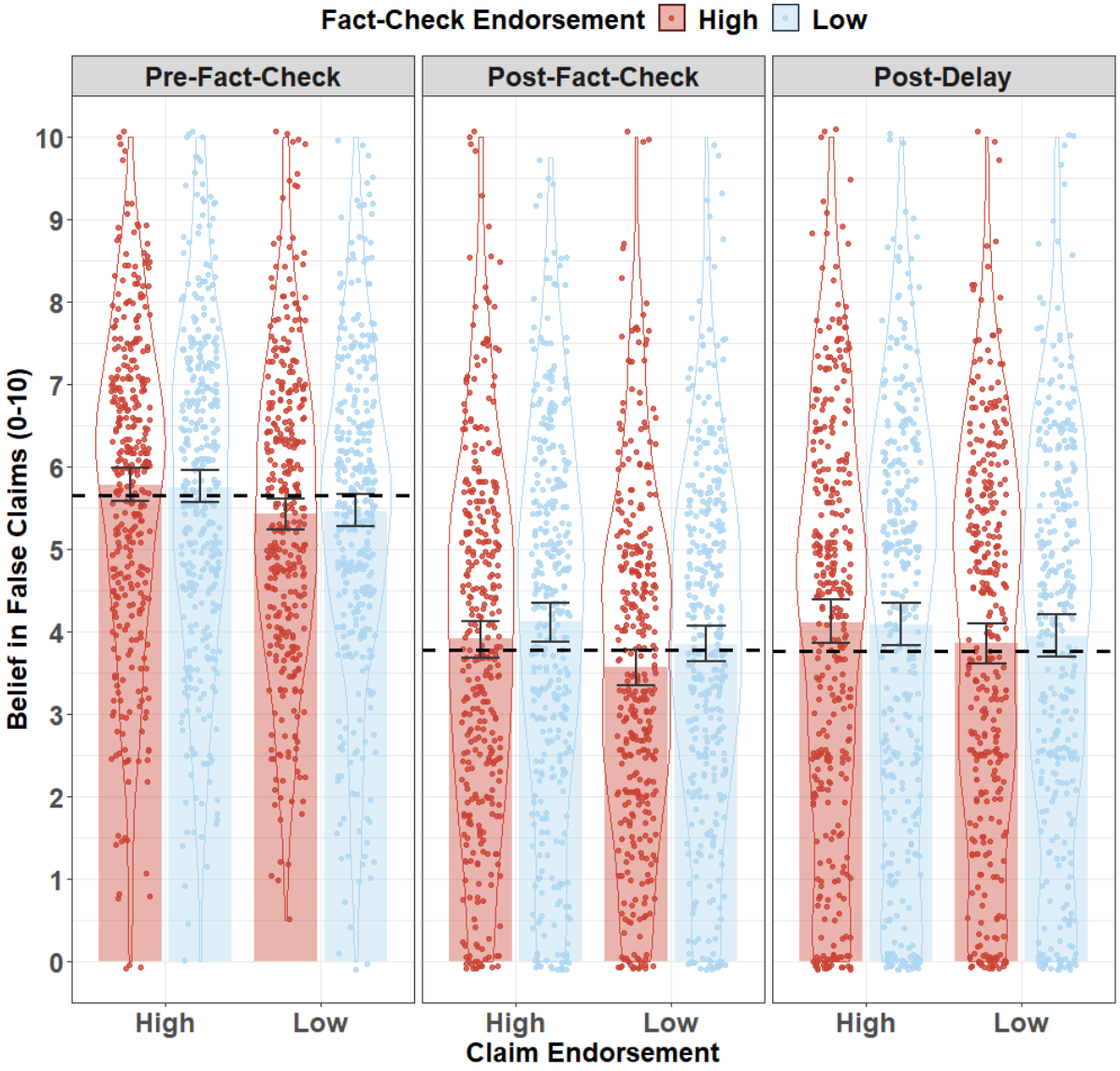
We first tested the effects of claim endorsement and fact-check endorsement at time 1 (pre-fact-check), to assess the immediate effect of claim endorsements, and to ensure fact-

check endorsement conditions did not significantly differ prior to fact-check exposure. There was a fixed effect of claim endorsement, $\beta = .20$, $SE = .05$, $z = 3.64$, $p < .001$. As expected, there was no statistical evidence for an effect of fact-check endorsement, $\beta = -.01$, $SE = .05$, $z = -.20$, $p = .841$, or an interaction between claim endorsement and fact-check endorsement, $\beta = .03$, $SE = .09$, $z = .35$, $p = .723$. These results replicate Experiment 1. The effect of claim endorsement indicates that belief was greater in misinformation with high endorsement than misinformation with low endorsement.

To assess the effect of fact-check endorsement, claim belief was then assessed across time-points 1 and 2. In line with Experiment 1, there was statistical evidence of fixed effects of claim endorsement, $\beta = .21$, $SE = .07$, $z = 2.96$, $p = .003$, and time, $\beta = -1.18$, $SE = .04$, $z = -33.80$, $p < .001$, and a time × fact-check-endorsement interaction, $\beta = -.15$, $SE = .07$, $z = -2.21$, $p = .027$. The effect of claim endorsement mirrors that at time 1, and the effect of time indicates that belief in misinformation reduced following fact-check exposure. The interaction indicates that the effect of fact-check endorsement significantly increased post fact-check exposure. To quantify this significant increase in the effect of fact-check endorsement, we assessed results of fact-check endorsement only at time-point 2 (post-fact-check). Replicating Experiment 1, there was a marginally significant fixed effect of fact-check endorsement, $\beta = -.14$, $SE = .07$, $z = -2.01$, $p = .044$, with high fact-check endorsement associated with greater belief updating.

Figure 4

Mean False-Claim Belief Ratings Across Time-points and Conditions in Experiment 2



Note. The horizontal dashed lines represent control-condition means at each time-point. Error bars represent 95% confidence intervals.

To assess whether effects of claim or fact-check endorsement were maintained over time, claim belief was assessed across time-points 2 and 3. As in Experiment 1, there were fixed effects of claim endorsement, $\beta = .21$, $SE = .08$, $z = 2.70$, $p = .007$, and fact-check endorsement, $\beta = -.15$, $SE = .07$, $z = -2.19$, $p = .028$, and a time × fact-check-endorsement interaction, $\beta = .15$, $SE = .07$, $z = 2.14$, $p = .032$ (unlike Experiment 1, the time × claim-endorsement interaction was n.s., $\beta = -.08$, $SE = .07$, $z = -1.10$, $p = .270$). The significant

effect of claim endorsement again mirrors that at time 1.[3] The time by fact-check-endorsement interaction suggests that the influence of fact-check endorsement significantly reduced over time. To quantify this reduction, we assessed the impact of fact-check endorsement on claim belief at time 3. As in Experiment 1, there was no statistical evidence of an effect of fact-check endorsement at time 3, $\beta < -.01$, $SE = .06$, $z = -.01$, $p = .992$.

Inference Scores. Mean inference scores across conditions are presented in Figure 5. There was no statistical evidence of an effect of claim endorsement, $\beta = .08$, $SE = .05$, $z = 1.46$, $p = .144$, fact-check endorsement, $\beta = .03$, $SE = .05$, $z = .62$, $p = .534$, or a claim-endorsement × fact-check-endorsement interaction, $\beta = .14$, $SE = .11$, $z = 1.28$, $p = .200$. Thus, the significant effect of claim endorsement on false-claim-congruent reasoning observed in Experiment 1 was not replicated in Experiment 2.

[3] Breaking this down further despite a nonsignificant time by claim-endorsement interaction—to keep analysis analogous to the pre-registered ANOVA analysis plan—the fixed effect of claim endorsement remained significant at time 2 but was nonsignificant at time 3. Note that the claim-endorsement effect at time 3 was significant in the ANOVA reported in the Supplement; the CLMM null effect was likely driven by relatively large between-subjects variability in slopes.

Figure 5

Mean False-Claim Inference Scores Across Conditions in Experiment 2

Note. The horizontal dashed lines represent control-condition means at each time-point. Error bars represent 95% confidence intervals.

Contrast to No-Endorsement Control

False-Claim Beliefs. To assess whether the effect of social endorsement on claim belief was driven by high claim (fact-check) endorsements increasing belief (belief updating), low claim (fact-check) endorsements decreasing belief (belief updating), or a combination of both, CLMM analyses were run contrasting false-claim belief in the high and low claim-endorsement and fact-check-endorsement conditions against the no-endorsement control condition. We first assessed effects of high claim endorsement (vs. no endorsement)

separately at times 1, 2, and 3. There was no statistical evidence of an effect of high claim endorsement on false claim beliefs at time 1, β = .03, *SE* = .09, *z* = .36, *p* = .723, time 2, β = .20, *SE* = .14, *z* = 1.50, *p* = .134, or time 3, β = .29, *SE* = .15, *z* = 1.90, *p* = .057.

We then assessed effects of low claim endorsement (vs. no endorsement) separately at time-points 1, 2, and 3. There was evidence of an effect of low claim endorsement at time 1, β = −.18, *SE* = .08, *z* = −2.20, *p* = .028, indicating false-claim belief in the low-claim-endorsement condition was significantly lower than in the no-endorsement control condition.[4] There was no statistical evidence of an effect of low claim endorsement at time 2, β = −.05, *SE* = .13, *z* = −.35, *p* = .723, or time 3, β = .14, *SE* = .15, *z* = .96, *p* = .338.

Next, we assessed the effects of high and low fact-check endorsement against no-endorsement control, separately at time-points 1, 2, and 3. There was no evidence of an effect of high fact-check endorsement on false claim beliefs at time 1, β = −.07, *SE* = .08, *z* = −.88, *p* = .378, time 2, β < .001, *SE* = .13, *z* < .01, *p* = .995, or time 3, β = .21, *SE* = .15, *z* = 1.44, *p* = .151, nor was there evidence of an effect of low fact-check endorsement at time 1, β = −.08, *SE* = .08, *z* = −.95, *p* = .342, time 2, β = .16, *SE* = .13, *z* = 1.21, *p* = .227, or time 3, β = .22, *SE* = .15, *z* = 1.51, *p* = .130.

True Claims

Again, analyses of true claims were exploratory; results are briefly described below, see Supplement B for details. There was no effect of claim endorsement at any time-point. There was a significant fact-check endorsement by time interaction between times 1 and 2, with greater fact-check endorsement associated with greater belief updating. There was also a fixed effect of fact-check endorsement across times 2 and 3, suggesting the effect of fact-check endorsement was maintained over time (unlike in Experiment 1, there was no time by fact-check endorsement interaction). True-claim inference scores were again analysed in an

[4] This effect was nonsignificant in an ANOVA follow-up contrast.

equivalent manner to the false claims. There was no statistical evidence of an effect of claim endorsement, fact-check endorsement, or a claim by fact-check endorsement interaction. Finally, there was no meaningful difference in true-claim belief or inference scores between the no-endorsement control condition and any of the endorsement conditions.

Discussion

Experiment 2 replicated the two primary effects observed in Experiment 1. First, there was again an effect of false-claim endorsement on misinformation belief, with high endorsement associated with greater misinformation belief. Second, there was again a significant influence of fact-check endorsement on claim belief at time 2. Experiment 2 therefore corroborates the conclusion that social-endorsement information can affect false-claim beliefs and belief updating. Note, however, that the effects of claim and fact-check endorsement were nonsignificant after a delay, at time-point 3. There was therefore mixed evidence for an enduring effect of social endorsements on beliefs and belief updating.

Inference-score results from Experiment 2 somewhat deviated from those of Experiment 1. Whereas in Experiment 1, high fact-check endorsement was associated with lower misinformation reliance than low fact-check endorsement, Experiment 2 found no significant difference. This suggests that endorsement effects may either (1) impact specific beliefs without influencing claim-related reasoning more broadly, or (2) have an immediate impact that reduces quickly over time.

The second aim of Experiment 2 was to investigate whether the observed differences between high- and low-endorsement conditions were driven by the impact of the high or low levels of endorsement, relative to a no-endorsement baseline, or both. There was some evidence that the effect of claim endorsement at time 1 was driven by low endorsement decreasing belief in false claims. However, there was no statistical evidence of a difference

between no-endorsement and either low- or high-endorsement conditions at any other time-point. As such, the overall evidence regarding the locus of the effect is inconclusive.

General Discussion

The current study assessed the role of social endorsement of false claims and fact-checks on misinformation belief and belief updating. It was hypothesised that belief in misinformation would be greater when it is associated with a high (vs. low) level of social endorsement. It was also hypothesised that corrections associated with a high (vs. low) level of endorsement would yield stronger belief updating (i.e., a reduced CIE).

In line with the first hypothesis, both experiments found that high false-claim endorsement was associated with greater false-claim belief both pre- and post-fact-check. This is consistent with the wider literature on the role of normative influence, in particular the bandwagon heuristic, in information evaluations (Metzger & Flanagin, 2013; Sundar, 2008). These findings serve as a proof-of-concept, and highlight the need to expand models of misinformation belief and continued influence, to more thoroughly consider the role of social factors (also see Ecker et al., 2022; Ecker, Sanderson et al., 2022; Vlasceanu & Coman, 2021). We note that in Experiment 2, the effect of claim endorsement on misinformation belief reduced to nonsignificant after a delay, suggesting that the role of claim endorsement on long-term beliefs may be limited. As such, although the current study demonstrates that false-claim endorsement influences misinformation belief immediately, further research is required to assess whether, and under what circumstances, this influence endures.

The social endorsement of corrections also influenced misinformation belief; in line with the second hypothesis, higher endorsement of the fact-check was associated with greater belief updating immediately. We note that this effect was not significant after the delay in both experiments, suggesting the effect of fact-check endorsement may be limited to immediate information appraisal. However, although there was no sustained effect of fact-

check endorsement on misinformation belief, the inferential reasoning results from Experiment 1 provide some evidence of an enduring effect of fact-check endorsement on reasoning, though this was not replicated in Experiment 2. As such, although the current study provides evidence of an immediate effect of fact-check endorsement on belief updating, we draw no strong conclusions regarding an enduring effect of fact-check endorsement.

Though the immediate effect of fact-check endorsement is in line with predictions and past research (specifically, Vlasceanu & Coman, 2021), the results of the current study suggest the effect of endorsement may be somewhat weaker than previously proposed. With this in mind it is important to note two things. Firstly, the level of engagement used in the high-endorsement conditions of the present study differed by an order of magnitude from the level used by Vlasceanu and Coman (approx. 1,000 likes vs. 10,000-30,000 likes, retweets, and comments). As such, it is possible that the level of social endorsement required to have an enduring influence on the effectiveness of fact-checks is higher than the level used in the current study. Secondly, the immediate effect of fact-check (as well as claim) endorsements should not be entirely overlooked. People are often overloaded with information on social media, and so decisions about information propagation behaviour are made quickly and without much deliberation (Bago et al., 2020; Van Bavel et al., 2021). As such, if endorsements influence sharing through a short-term influence on belief, this may have a meaningful impact even in the absence of a persistent effect on belief (Avram et al., 2020).

Unsurprisingly, the largest effect on belief observed in the current study was driven by the provision of corrections: Across the board misinformation that had been corrected was believed substantially less than misinformation that had not been corrected. This reiterates that although factors associated with corrections can enhance or reduce their effectiveness, providing a correction to misinformation is often far more important when it comes to belief updating than other factors (Swire-Thompson et al., 2021). In fact, the effect of information

endorsement being comparatively small was entirely expected given it is a relatively less important factor, especially when it comes to the correction. Specifically, in the present study, when participants encountered corrections, they had already formed a level of belief in the misinformation, which may have served as an anchor during correction processing (Hogarth & Einhorn, 1992; Lieder et al., 2018). Moreover, the most salient information provided by the fact-check is whether the initial claim was corrected or affirmed, as opposed to the fact-check's level of endorsement. Accordingly, the fact-check's endorsement level carries less persuasive utility compared to the correction or affirmation, and participants' level of post-fact-check belief will have likely been influenced primarily by their level of pre-fact-check belief and the corrective-versus-affirmative nature of the fact-check per se. Taking these factors into consideration, the fact that the level of fact-check endorsement had a replicable effect on immediate belief updating is noteworthy.

From an applied perspective, the current study suggests that people's immediate susceptibility to misinformation may be particularly high for misinformation with high levels of endorsement, as is typical for posts that go viral on social media. Though the effects observed in the current study were small, small shifts in belief have the capacity to have a meaningful impact at scale and over repeated exposures. This is of concern as: (1) dangerous online misinformation often has characteristics that make it more likely than true information to go viral (King & Wang, 2021), and (2) people are more likely to be exposed to misinformation, potentially multiple times (Fazio, 2020), if it is viral (Del Vicario et al., 2016). Additionally, as a number of platforms (e.g., Twitter, Instagram) only allow for endorsement (and not *dis*-endorsement, i.e., "dislikes") via engagement metrics, it is possible that people overestimate the extent to which endorsement information represents normative belief.

One potential limitation of the current study relates to the experimental design. Specifically, people only received social information through quantity of likes, the misinformation was non-controversial, and participants were required to read each post they were exposed to. These design features were necessary to minimise potential confounds and assess the endorsement effects; however, they may have reduced external validity. That is, participants may have placed more weight on the number of likes than they typically would, especially given they were unfamiliar with the information sources and were unlikely to have strong opinions about the claims. That said, it is possible that engagement information interacts with other social or worldview factors, potentially enhancing its role in belief formation and updating compared to the level observed in the current study (as was seen in Borah & Xiao, 2018).

We also recognise that the presentation of fact-checks in close temporal proximity to the corresponding claims provided optimal circumstances to observe endorsement effects. Moreover, the within-subjects design, which was chosen because it allowed us to control for variability in how participants responded to the key manipulation, provided reasonable statistical power. Thus, the effects may be smaller or absent under different conditions. However, the main aim of this study was to provide proof-of-concept that even relatively limited social information may influence how people appraise misinformation and corrections; as such, creating conducive conditions was entirely intentional. We also note that presentation conditions were not dissimilar to how people routinely encounter misinformation and corrections on social media, with fact-checks often appearing in the comment section of posts.

In conclusion, we provide evidence that the perceived normative nature of misinformation, indicated by engagement metrics, directly influences its believability. With society's increasing reliance on unregulated social-media platforms for information

acquisition, this influence of endorsement information may be reason for concern, particularly given that engagement information can be maliciously manipulated (Shao et al., 2018), and misleading and conspiratorial claims often have characteristics designed to enhance their level of endorsement (King & Wang, 2021). Although endorsement of corrections had a small influence on participants' misinformation belief immediately, an enduring effect on misinformation belief could not be statistically substantiated. Although further research is required to assess the impact this immediate effect could have on factors such as misinformation propagation, the current pattern of results suggest that social endorsement of claims and fact-checks can influence people's initial belief in misinformation, as well as their immediate belief updating in the face of corrections.

References

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, *70*(9), 1–70. https://doi.org/10.1037/h0093718

Avram, M., Micallef, N., Patil, S., & Menczer, F. (2020). Exposure to social engagement metrics increases vulnerability to misinformation. *The Harvard Kennedy School Misinfromation Review*, *1*(5). https://doi.org/10.37016/mr-2020-033

Bago, B., Rand, D. G., & Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of Experimental Psychology: General*, *149*(8), 1608–1613. https://doi.org/10.1037/xge0000729

Bode, L., & Vraga, E. K. (2018). See something, say something: Correction of global health misinformation on social media. *Health Communication*, *33*(9), 1131–1140. https://doi.org/10.1080/10410236.2017.1331312

Borah, P., & Xiao, X. (2018). The importance of 'likes': The interplay of message framing, source, and social endorsement on credibility perceptions of health information on Facebook. *Journal of Health Communication*, *23*(4), 399–411. https://doi.org/10.1080/10810730.2018.1455770

Brydges, C. R., Gignac, G. E., & Ecker, U. K. H. (2018). Working memory capacity, short-term memory capacity, and the continued influence effect: A latent-variable analysis. *Intelligence*, *69*, 117–122. https://doi.org/10.1016/j.intell.2018.03.009

Butler, L. H., Fay, N., & Ecker, U. K. H. (2022). Social endorsement influences the continued belief in corrected misinformation. Retrieved from https://osf.io/vt7p6/.

Christensen, R. H. B. (2018). Cumulative link models for ordinal regression with the R Package ordinal. *Unsubmitted Manuscript*, 1–40. www.jstatsoft.org/

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity.

Annual Review of Psychology, *55*, 591–621. https://doi.org/10.1146/annurev.psych.55.090902.142015

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., & Caldarelli, G. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, *113*(3), 554–559. https://doi.org/10.1073/pnas.1517441113

Donzelli, G., Palomba, G., Federigi, I., Aquino, F., Cioni, L., Verani, M., Carducci, A., & Lopalco, P. (2018). Misinformation on vaccination: A quantitative analysis of YouTube videos. *Human Vaccines and Immunotherapeutics*, *14*(7), 1654–1659. https://doi.org/10.1080/21645515.2018.1454572

Ecker, U. K. H., & Antonio, L. M. (2021). Can you believe it? An investigation into the impact of retraction source credibility on the continued influence effect. *Memory & Cognition*, *49*(4), 631–644. https://doi.org/10.3758/s13421-020-01129-y

Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, *1*(1), 13–29. https://doi.org/10.1038/s44159-021-00006-y

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Fazio, L. K. (2020). Repetition increases perceived truth even for known falsehoods. *Collabra: Psychology*, *6*(1). https://doi.org/10.1525/collabra.347

Flanagin, A. J. (2017). Online social influence and the convergence of mass and interpersonal communication. *Human Communication Research*, *43*(4), 450–463. https://doi.org/10.1111/hcre.12116

Guillory, J. J., & Geraci, L. (2013). Correcting erroneous inferences in memory: The role of

source credibility. *Journal of Applied Research in Memory and Cognition*, *2*(4), 201–209. https://doi.org/10.1016/j.jarmac.2013.10.001

Hornsey, M. J., Majkut, L., Terry, D. J., & McKimmie, B. M. (2003). On being loud and proud: Non-conformity and counter-conformity to group norms. *British Journal of Social Psychology*, *42*(3), 319–335. https://doi.org/10.1348/014466603322438189

Jia, L., Shan, J., Xu, G., & Jin, H. (2020). Influence of individual differences in working memory on the continued influence effect of misinformation. *Journal of Cognitive Psychology*, *32*(5–6), 494–505. https://doi.org/10.1080/20445911.2020.1800019

Johnson, H. M., & Seifert, C. M. (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

Kaplan, M. F., & Miller, C. E. (1987). Group decision making and normative versus informational influence: Effects of type of issue and assigned decision rule. *Journal of Personality and Social Psychology*, *53*(2), 306–313. https://doi.org/10.1037/0022-3514.53.2.306

Kendal, R. L., Boogert, N. J., Rendell, L., Laland, K. N., Webster, M., & Jones, P. L. (2018). Social learning strategies: Bridge-building between fields. *Trends in Cognitive Sciences*, *22*(7), 651–665. https://doi.org/10.1016/j.tics.2018.04.003

Kim, J. (2021). The meaning of numbers: Effect of social media engagement metrics in risk communication. *Communication Studies*, *72*(2), 195–213. https://doi.org/10.1080/10510974.2020.1819842

King, K. K., & Wang, B. (2021). Diffusion of real versus misinformation during a crisis event: A big data-driven approach. *International Journal of Information Management*, 102390. https://doi.org/10.1016/j.ijinfomgt.2021.102390

Lewandowsky, S., Cook, J., Ecker, U. K. H., Albarracín, D., Amazeen, M. A., Kendeou, P., Lombardi, D., Newman, E. J., Pennycook, G., Porter, E., Rand, D. G., Rapp, D. N., Reifler, J., Roozenbeek, J., Schmid, P., Seifert, C. M., Sinatra, G. M., Swire-Thompson, B., van der Linden, S., … Zaragoza, M. S. (2020). *The debunking handbook 2020*. https://doi.org/10.17910/b7.1182

McElreath, R. (2020). Statistical rethinking: A bayesian course with examples in R and stan. In *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC Press. https://doi.org/10.1201/9781315372495

Mena, P., Barbe, D., & Chan-Olmsted, S. (2020). Misinformation on Instagram: The impact of trusted endorsements on message credibility. *Social Media + Society*, *6*(2), 205630512093510. https://doi.org/10.1177/2056305120935102

Metzger, M. J., & Flanagin, A. J. (2013). Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of Pragmatics*, *59*, 210–220. https://doi.org/10.1016/j.pragma.2013.07.012

Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. *Journal of Communication*, *60*(3), 413–439. https://doi.org/10.1111/j.1460-2466.2010.01488.x

Peters, K., Chen, Y., Kaplan, A. M., Ognibeni, B., & Pauwels, K. (2013). Social media metrics - A framework and guidelines for managing social media. *Journal of Interactive Marketing*, *27*(4), 281–298. https://doi.org/10.1016/j.intmar.2013.09.007

Qiu, X., Oliveira, D. F. M., Sahami Shirazi, A., Flammini, A., & Menczer, F. (2017). Limited individual attention and online virality of low-quality information. *Nature Human Behaviour*, *1*(7), 1–15. https://doi.org/10.1038/s41562-017-0132

Shao, C., Ciampaglia, G. L., Varol, O., Yang, K. C., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature Communications*, *9*(1), 1–9.

https://doi.org/10.1038/s41467-018-06930-7

Smith, E. R., & Semin, G. R. (2007). Situated social cognition. *Current Directions in Psychological Science*, *16*(3), 132–135. https://doi.org/10.1111/j.1467-8721.2007.00490.x

Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. *Digital Media, Youth, and Credibility*, 73–100. https://doi.org/10.1162/dmal.9780262562324.073

Swire-Thompson, B., Cook, J., Butler, L. H., Sanderson, J. A., Lewandowsky, S., & Ecker, U. K. H. (2021). Correction format has a limited role when debunking misinformation. *Cognitive Research: Principles and Implications*, *6*(1), 1–15. https://doi.org/10.1186/s41235-021-00346-6

Van Bavel, J. J., Harris, E. A., Pärnamets, P., Rathje, S., Doell, K. C., & Tucker, J. A. (2021). Political psychology in the digital (mis)information age: A model of news belief and sharing. *Social Issues and Policy Review*, *15*(1), 84–113. https://doi.org/10.1111/sipr.12077

Vlasceanu, M., & Coman, A. (2021). The impact of social norms on health-related belief update. *Applied Psychology: Health and Well-Being*, 85–91. https://doi.org/10.1111/aphw.12313

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

Walter, N., & Tukachinsky, R. (2020). A meta-analytic examination of the continued influence of misinformation in the face of correction: How powerful is it, why does it happen, and how to stop it? *Communication Research*, *47*(2), 155–177. https://doi.org/10.1177/0093650219854600