Negative news dominates fast and slow brain responses and social judgments even after source credibility evaluation

Julia Baum a, b, *, Rasha Abdel Rahman a, b, *

a Humboldt-Universität zu Berlin, Faculty of Philosophy, Berlin School of Mind and Brain, Germany
b Humboldt-Universität zu Berlin, Faculty of Life Sciences, Department of Psychology, Germany

A R T I C L E   I N F O

Keywords:
Emotional person knowledge
Cognitive intervention
Misinformation
Event-related potentials
Evaluative learning
Social judgment
Source credibility

A B S T R A C T

Remedies to counter the impact of misinformation are in high demand, but little is known about the neurocognitive consequences of untrustworthy information and how they can be mitigated. In this preregistered study, we investigated the effects of social-emotional headline contents on social judgments and brain responses and whether they can be modulated by explicit evaluations of the trustworthiness of the media source. Participants (N = 30) evaluated—and clearly discerned—the trustworthiness of news sources before they were exposed to person-related news headlines. Despite this intervention, social judgments and brain responses were dominated largely by emotional headline contents. Results suggest differential effects of source credibility might depend on headline valence. Electrophysiological indexes of fast emotional and arousal-related brain responses, as well as correlates of slow evaluative processing were enhanced for persons associated with positive headline contents from trusted sources, but not when positive headlines stemmed from distrusted sources. In contrast, negative headlines dominated fast and slow brain responses unaffected by explicit source credibility evaluations. These results provide novel insights into the brain mechanisms underlying the “success” of emotional news from untrustworthy sources, suggesting a pronounced susceptibility to negative information even from distrusted sources that is reduced for positive contents. The differential pattern of responses to misinformation in mind and brain sheds light on the cognitive mechanisms underlying the processing of misinformation and possible strategies to avoid their potentially detrimental effects.

Facing today’s massive online information overload, it is especially important for us to distinguish trustworthy and well-vetted news from all kinds of rumours, misinformation or lies. But how can we guard ourselves against false news biasing our cognitive processing and social judgments? Previous research on this topic using large-scale behavioural online studies has revealed how well we can discern the trustworthiness of news and how they influence our online behavior, such as whether we share information or not (e.g., Pennycook et al., 2021; Pennycook and Rand, 2019). Our study aims at complementing this research by shedding light on the cognitive and brain mechanisms underlying the processing of news contents of varying levels of trustworthiness. Neurocognitive information is thus far scant (pointed out also by e.g., Lazer et al., 2018; Vosoughi et al., 2018), but may provide valuable cues helping us to better understand our susceptibility to misinformation and, in perspective, to identify potential cognitive strategies as a protection against misinformation and false news.

In the present study, we focus on the role of emotional contents of misinformation, assuming that emotion may be an important ingredient for the “success” of misinformation (Baum et al., 2018; Baum and Abdel Rahman, 2020; Baum et al., 2020; Martel et al., 2020). Specifically, we expose participants to negative, positive and comparatively neutral person-related news headlines from trusted and distrusted media sources, investigating the subsequent, and likely memory-related, neural and social consequences of such news. We analyze the neurocognitive signatures and social judgments as well as the confidence with which these judgments are made in response to what the news claim, where social judgments are valence ratings that refer to the person and their social characteristics. As a crucial manipulation we test if leveraging people’s ability to explicitly discern the credibility of sources can serve as a cognitive guard against the influences of misinformation. This minimal cognitive intervention was implemented by deliberately evaluating the trustworthiness of media sources before reading their headlines.

We routinely form social opinions from second-hand information, and even untrustworthy emotional information can bias our judgments (Baum et al., 2018; Baum and Abdel Rahman, 2020; Baum et al., 2020). Without first-hand experience, we can learn the affective value of others through even sparse declarative, verbally transmitted person-related...
information via mechanisms of evaluative learning (for example, “He bullied his apprentice”; Bliss-Moreau et al., 2008; Ferrari et al., 2020; Mattarozzi et al., 2014; Todorov and Olson, 2008). The robust consequences of the affective information are reflected in early and later neural signatures of face processing and person evaluation, in modulations of the access to visual consciousness as well as explicit social judgments (Anderson et al., 2011, but see for conflicting findings Rabovsky et al., 2016 and Stein et al., 2017; Aviezer et al., 2017; Abdel Rahman, 2011; Baum and Abdel Rahman, 2020; Baum et al., 2018; Eiserbeck and Abdel Rahman, 2020; Eiserbeck et al., 2021; Galli et al., 2006; Junghöfer et al., 2016; Kessler and Strehlow, 2017; Luo et al., 2016; Suess et al., 2014; Wieser et al., 2014; Xu et al., 2016).

False news and misinformation with respect to social-emotional information can have detrimental effects. Therefore, it is paramount to distinguish trustworthy from untrustworthy information. The credibility of the source is an important index and simple heuristic to assess information veracity. However, knowing and understanding media source trustworthiness does not seem to automatically guard against the influences of emotional contents of news, although people have distinct opinions on the credibility of media sources and acknowledge contextual source information when they are confronted with news (Baum and Abdel Rahman, 2020; see also Metzger and Flanagan, 2013; Pennycook and Rand, 2019). Regardless of perceived source credibility, negative and positive emotional contents may lead to strongly valenced and confident social judgments of persons associated with such headlines (Baum and Abdel Rahman, 2020; see also Baum et al., 2020). Here, we investigate how susceptible these effects are to a simple cognitive intervention aimed at strengthening the ability to consider source credibility. We test if explicitly attending to and evaluating media sources’ trustworthiness before being confronted with their news headlines can be used as a protection against the effects of untrustworthy headlines (Fig. 1).

We base our predictions on effects of the valence and credibility of information on the class of dual-process theories and appraisal theories of emotion that differentiate between fast, relatively automatic and slow, more controlled information processing (Cunningham and Zelazo, 2007; FeldmanHall and Shenhav, 2019; Gawronski and Bodenhausen, 2006; Kahneaman, 2003; Lieberman, 2007; Strack and Deutsch, 2004; Yonelinas, 2002; and Ellsworth and Scherer, 2003; Scherer, 2001). Fast processing should primarily be affected by emotional salience and arousal of the information, whereas slow processing should be sensitive to the meaning or relevance of emotion in a given context, such as a current task, goal, or reappraisal and should therefore be sensitive to the trustworthiness of the information, resulting in leveraged judgments. Early and late processing is investigated with event-related potentials (ERPs) as high-temporally resolving signatures of brain activity allowing us to track the processing of social-emotional information as it unfolds over time. Specifically, relatively fast processing related to reflexive and arousal-related emotional processes is indexed by the early posterior negativity (EPN; occurring about 200–300 ms after stimulus onset at occipito-temporal brain regions; e.g., Junghöfer et al., 2001; Schupp et al., 2003, 2004). Slower, more controlled elaborate and reflective processing is associated with the late positive potential (LPP; about 400–600 ms at centro-parietal regions; e.g. Sabatinielli et al., 2013; Schacht and Sommer, 2009a; Schupp et al., 2004). The EPN has been found to be primarily affected by emotional salience and arousal whereas effects on the LPP were shown to be sensitive to the meaning or relevance of emotion in a given context, such as a current task, goal, or reappraisal (Blechert et al., 2012; Herbert et al., 2011; 2013; Rollecke et al., 2012; Schacht and Sommer, 2009b; Schindler et al., 2019, 2020, 2021).

Concerning our minimal cognitive intervention of actively discerning the credibility of the media sources, we assume that this should enhance the saliency and availability of the source with its perceived trustworthiness during later processing of the headlines. As a result, evaluative processing and social judgments should be more strongly affected by headlines form explicitly trusted sources, whereas emotional headlines from explicitly distrusted sources should be put into perspective, resulting in weaker or no effects (Fig. 1, Phase 2). The current implementation of such a cognitive guard in the context of misinformation can be related to the concept of nudging or boosting that have recently been shown to enhance discernment of news accuracy (Pennycook et al., 2020, 2021; Roozenbeek and van der Linden, 2019) and have been proposed as a protection against misinformation (Kozyreva et al., 2020; Lorenz-Spreen et al., 2020). Nudges offer cues that steer people’s attention and boosts encourage people’s cognitive competencies to enable better decisions that align with their goals (Hertwig & Grüne-Yanoff, 2017). Here, we apply one aspect of these concepts by steering attention towards the news’ trustworthiness during processing of emotional headlines. To our knowledge, the scope and limits of such a cognitive intervention have not yet been investigated in the context of the effects of social-emotional information and its underlying neurocognitive mechanisms.

As preregistered (osf.io/hqwy2) we investigated the effects of emotional news from trusted and distrusted media sources with social judgments and ERPs to reveal the underlying brain responses at different levels of processing. Fast emotional brain responses in the EPN and slower evaluation-related responses in the LPP are expected to be dominated by negative and positive contents, regardless of source credibility (Baum and Abdel Rahman, 2020). We hypothesized that if this cognitive intervention is effective, fast emotion effects induced by headline contents in the EPN may still occur independent of source credibility, whereas emotion effects should be modulated by source credibility in the LPP, reflecting more controlled evaluative processing that may also take context information into account. While enhanced processing of emotional contents is expected for trusted sources, emotion effects of headlines from distrusted sources should be reduced or absent in the LPP, as well as in the behavioral measures of social judgments and the confidence in the judgments. Furthermore, the success of the cognitive intervention may depend on the valence of information, such that source credibility may primarily modulate the influence of positive, but not negative information (as hypothesized also in Baum et al., 2018; Baum and Abdel Rahman, 2020). This expectation is based on the presumed prioritization of negative information as a protection against potential threat (e.g., Öhman & Mineka, 2001). Alternatively, if the cognitive intervention is not effective, we expected emotional headline contents to strongly affect information processing and social judgments irrespective of source credibility (Baum and Abdel Rahman, 2020).

1. Method

1.1. Participants

The sample size was planned according to the requirements of the counterbalancing and based on power analyses preregistered on OSF (osf.io/hqwy2), see Supplemental Information (SI) page 1 and SI-tables S1, S2 (the SI is linked in the end of the article). The sample consisted of 30 participants (mean age = 25 (SD = 5.36), 25 females, 5 males, 90% right-handed). Nine participants were excluded and replaced with new participants (reasons: excess noise in the EEG, being familiar with face databases, insufficient eyewitness, not following task instructions). Participants were compensated in form of course credit or money. They were (de)briefed about the procedures and signed informed consent. The study was approved by the local ethics committee.

1.2. Materials

Materials were identical to previous experiments (Baum and Abdel Rahman, 2020; Baum et al., 2020). Face stimuli were 24 unfamiliar faces exhibiting neutral facial expressions (various databases, e.g. Ebner et al., 2016; Langner et al., 2010). Faces were shown as colored headshots on gray background in the person likability ratings in Phase 1 and the social judgment tasks in Phase 2 (2.7 x 3.5 cm, viewing distance 70 cm). Eight
Fig. 1. Overview of the study design with two phases (Phase 1 is shown here) and key experimental manipulations. A. Depicted is the procedure of Phase 1 with the cognitive intervention in form of source classifications, the news exposure, and framing pre- and post-exposure person likability ratings as manipulation check. B. – D. The cognitive intervention before news exposure entailed media source classifications according to their perceived familiarity, trustworthiness, and likability. Sources differed on trustworthiness (C.) and likability (D.). Sources used in the experiment are existing and widely distributed mainstream German news media (e.g., Tagesschau and Bild) that were selected based on pre-ratings of high or poor credibility. Similar English-speaking media outlets are for example The New York Times, BBC News, Fox News or The Sun. E. During news exposure, participants were exposed to authentic websites of these media outlets containing the prominent source logo, the face, and the headline with all other details remaining blurred. Unfamiliar faces were presented on trusted or distrusted media websites with positive, neutral, or negative person-related headlines. The assignment of source, face, and headline was counterbalanced across participants. F. Participants rated the likability of persons before and after news exposure as manipulation check. Headline contents modulated person likability independent of source credibility.
well-known filler faces were added to make the overall existence of the persons credible.

For the classifications of sources in Phase 1, colored screenshots of the eight media outlet’s logos were presented in similar size (2.7 × 3.5 cm). We chose news media sources used in this study according to pre-ratings of credibility and familiarity by a different sample of German participants (please see SI-page 11 and Baum and Abdel Rahman, 2020).

For the news exposure in Phase 1, static experimental websites with the original online layouts of well-known German news media outlets were displayed. Each display entailed the source logo, headline, and face with all other details blurred (for examples, see Fig. 1, Phase 1). Headlines were of either neutral, positive, or negative valanced content (e.g., neutral: Startup scene: He invests in smart household appliances; positive: Scientific breakthrough: This researcher restores sight for blind people; negative: Berlin: This casino owner forced tiener to smuggle cocaine; all headlines can be found in the SI, page 12–14). Valence and arousal were confirmed by pre-ratings, further showing that positive and negative headline content were equally more arousing than neutral (see SI-page 11, and Baum and Abdel Rahman, 2020). The assignment of the unfamiliar faces was equally counterbalanced to headlines (neutral, negative, and positive) and to sources (trusted, distrusted).

1.3. Procedure

The procedure was mainly equivalent to previous experiments, with the exception that we introduced a cognitive intervention before news exposure (Baum and Abdel Rahman, 2020; Baum et al., 2020) and based on a well-established design, cf. Abdel Rahman, 2011; Baum et al., 2018; Suess et al., 2014). The experiment consisted of two phases (Fig. 1). In Phase 1, participants first rated all persons on likability. We included this likability rating as it can be done based on a spontaneous feeling of liking without further information about a person and thereby provides a global pre-post measure as manipulation check. As a cognitive intervention they then rated all media sources on their familiarity, credibility, and likability. As instruction participants were told that they will receive media reports about the persons and that first they should answer some questions. The news exposure followed, where in each trial a website was presented completely blurred except for the source logo for one second followed by five seconds of unblurred logo, face, and headline. Websites were presented in blocks of eight, each block including all experimental conditions and two fillers and repeated five times in total across news exposure. We previously demonstrated via eye-tracking that participants look at the media source logo during news exposure (Baum and Abdel Rahman, 2020). Therefore, we conclude that the information about the source is available during news exposure. Participants were kept engaged during news exposure by occasionally answering short yes-or-no questions about the persons (e.g. Is the behavior of this person common?, asked in about 20% of the trials of Phase 1). After news exposure they took a 15-minute break and ended Phase 1 with a post-exposure person likability rating.

Phase 2 started with the social judgment task as main task while the EEG was acquired. Each face was presented in isolation and participants judged on a 5-point scale how positive, neutral, or negative the depicted person was based on the information of Phase 1. For EEG data quality, the task was repeated 20 times block-wise (resulting in 80 trials per condition) and participants were told that this was necessary for EEG measurements. In each trial, a fixation cross was presented for 0.5 s, a face was presented until response or maximally 3.5 s followed by a 0.5 s fixation cross as inter-trial-interval. As a secondary task, the confidence in judgment task was conducted separately after the main task was completed. Participants first gave an overall social judgment of the person on the same scale as in the main task, however now this was in each trial directly followed by a rating of their confidence in this judgment on a slider scale ranging from low to high confidence (0–100, 1 increment; this procedure was adopted form the confidence literature, see Baum et al., 2020; Frömer et al., 2021).

After the main experiment, additional tests followed. In a source memory task the participants saw each face in isolation and were asked to decide in which of two depicted sources (one trusted, one distrusted) the face had appeared in Phase 1. If the face had appeared in a trusted source, the distractor was a randomly selected distrusted source and vice versa. Across all conditions, source memory was better than chance (M = 0.61, 95%-CI [.57, 0.65], t(29)= 5.18, p < .001, d = 5.2). We had no hypotheses on this task and included it as first step for future investigations into explicit retrieval processes, as specific source memory seems to be relatively weak (cf. Johnson et al., 1993). Importantly, eye-tracking evidence confirms for this experimental design that source information is sampled by the participants and available during news exposure in Phase 1 (Baum and Abdel Rahman, 2020). Next, participants rated all media sources on arousal. To measure source logo evoked ERPs, the rating was repeated 10 times (i.e., 40 trials per condition; ERPs evoked by source logos can be found in the SI, page 15). Trusted sources were rated as marginally more arousing than distrusted sources (b = 0.81, SE=.03, 95%-CI [.04, 1.57], t = 2.07, p=.046; SI-Table S18 and SI-Fig. S1), suggesting an at least initially increased arousal towards distrusted sources. However, please note that this effect is likely affected by the characteristic layouts that differ in perceptual salience from trusted sources (e.g., red font). The reversed effect in the rating may be due to later cognitive or strategic adjustments described above. At last, participants wrote down contents of each headline on a list of all faces to check that they broadly remembered the headlines.

The direction of scales was counterbalanced but consistent for one subject in all tasks. Therefore, there were two versions with version one having 5-point scales range from positive (left) to negative (right), and in version two from negative (left) to positive (right). The experiment lasted about 70 minutes and participants were compensated for all time spent at the lab.

1.4. EEG data recording and preprocessing

The EEG was recorded with BrainAmpDC amplifiers, from 62 Ag/AgCl-electrodes as specified by the extended 10–20 system, referenced to the left mastoid with FCz as Ground Electrode. Impedance was kept under 5kΩ. EEG data was recorded at a sampling rate of 5 kHz and down-sampled to 500 Hz using a low-cutoff of 0.016 Hz and a high-cutoff of 1000 Hz. Horizontal and vertical electrooculograms were obtained with peripheral electrodes at the left and right canthi of both eyes, and above and below the left eye. A short calibration procedure traced individual eye movements after the experiment, that were later used to correct for eye movement artifacts.

Offline, the continuous EEG was transformed to average reference and low-pass filtered at 30 Hz pass-band edge (zero-phase FIR-filter with transition band width of 7.5 Hz and cutoff frequency (–6 dB):33.75 Hz, EEGlab-toolbox version 13.5.4b; Delorme and Makeig, 2004). Using BESA (Berg & Scherg, 1991), we removed artifacts due to eye movements by applying a spatiotemporal dipole modeling procedure for each participant individually. Trials with remaining artifacts were rejected, i.e. trials with amplitudes over ±200μV, changing more than 50μV between samples or more than 200μV within single epochs, or containing baseline drifts. Error- and artifact-free EEG data was segmented into epochs of 1 s, starting 100 ms prior to stimulus onset, with a 100 ms pre-stimulus baseline. For EEG analysis, per participant an average of 73 trials per condition remained and in each condition 91% of trials were kept overall (neutral-trusted 2174, neutral-distrusted 2194, negative-
trusted 2189, negative-distrusted 2205, positive-trusted 2194, positive-distrusted 2196). Trials where no judgment was given, or reaction time was below 200 ms were excluded (in the social judgment task that were 133 out of 14,400).

1.5. Data analysis

ERP analyses focused on the EPN (at electrode sites PO7, PO8, P09, PO10, TP9, TP10, 200–350 ms after face stimulus onset) and the LPP component (Pz, CPz, POz, P3, P4, 400–600 ms; regions of interest were preregistered and equal to Baum and Abdel Rahman, 2020). Amplitudes were averaged over ROIs and time windows on single-trial level.

We performed linear mixed-effects regression models (LMMs) on single-trial data of behavioral measures and ERPs (Frömer et al., 2018; Bates et al., 2015b; lme4 v.1.1–17 and confirmed in v.1.1.26 in R). We tested the significance of fixed effects coefficients (p-value < 0.05) by Satterthwaite approximation (summary function of lmerTest v.3.0–1, Ruznêtsova et al., 2017). For each dependent variable, the model was specified with fixed effects for the experimental factors headline content (positive, negative, neutral; with neutral as the reference level) and source credibility (trusted, distrusted; with distrusted as the reference level) and their interaction. Both factors were modeled as repeated contrasts that compare the means of factor levels to the respective reference level. Model coefficients represent our hypotheses that expect emotion effects of negative vs. neutral and of positive vs. neutral headline content, each in interaction with source credibility, with reduced or absent effects of headline content for distrusted sources (see Schad et al., 2020 for details on testing a-priori hypotheses through contrast specification in LMMs). We fitted models with a maximal crossed random-effects structure correcting for by-subjects and by-face-stimuli random intercepts and slopes. If necessary, random-slopes correlation parameters were set to zero and slopes explaining zero variance were omitted to achieve convergence and avoid overparameterization (Bates et al., 2015a; final random structures are reported in the results Tables). To test our hypotheses that emotion effects may be present only for trusted but not distrusted sources, we tested emotion effects separately for each source credibility condition as a follow-up (via emmeans v.1.4.6, Lenth, 2020, with false-discovery-rate adjusted p-values, Benjamin and Hochberg, 1995). We report point estimates ($\hat{b}$), 95% confidence intervals for LMMs, standard errors, t-values, and p-values for the fixed effects coefficients. Data and code can be accessed online (osf.io/hqwy2)

2. Results

2.1. Phase 1

2.1.1. Classification of media sources as minimal cognitive intervention

Participants rated sources that were assigned to the distrusted condition based on pre-ratings as untrustworthy and less likable, while they rated sources of the trusted condition as trustworthy and likable (source credibility effect in trust ratings: $b = 2.67, SE = 0.27, 95%-CI [2.16, 3.19], t = 10.08, p < .001, and in likability ratings: $b = 2.11, SE = 0.33, 95%-CI [1.47, 2.75], t = 6.48, p < .001$; see Fig. 1 and SI-Tables S3, S4). Furthermore, both, distrusted and trusted sources were rated as highly familiar ($M = 0.92, 95%-CI [.86, 0.99]; no difference between source conditions, $F(1,29) = 0.0, p = 1$).

2.1.2. Person likability rating as manipulation check

Before participants were exposed to news, person likability ratings were on average neutral (SI-Table S5 to S7; see Figure 1 Panel F). Post-exposure ratings showed that persons associated with negative headlines were disliked and persons associated with positive headlines were liked relatively to neutral headlines ($b = -1.56, SE = 0.10, 95%-CI [-1.77, -1.36], t = -14.95, p < .001 and $b = 0.80, SE = 0.08, 95%-CI [.64, 0.97], t = 9.51, p < .001$, respectively). In contrast, person likability was not affected by source credibility (all t-values < 0.9, all p-values > 0.4).

2.2. Phase 2

2.2.1. Behavioral results in the social judgment task

In the social judgment task, participants judged each person presented in isolation based on the information from Phase 1 (see Figure 2). For associated negative headlines, persons were judged as more negative relative to neutral headlines, and judgments based on negative headlines were faster compared to neutral headlines (please see Table 1, for all descriptive results see SI-Table S8). In contrast to expectations of a successful intervention, source credibility did not modulate the negative headline effects in judgments and their latencies (Table 1). Social judgments based on negative vs. neutral headlines were more negative and faster for both, trusted and distrusted sources (Table 2).

For positive headlines, social judgments of associated persons were more positive and also faster compared to neutral headlines (Table 1). Social judgments were not modulated by source credibility (Table 2). While for latencies of judgments the interaction effect was not significant (Table 1), preregistered direct comparisons show that judgments were faster for positive headlines vs. neutral only in the case of trusted sources (Table 2).

Post-hoc (non-preregistered), we included repetition as a covariate to test whether social judgments and their latencies were affected by repeating the task, which was necessary to ensure EEG data quality. The three-way interactions were not significant (all t-values < 1.8, all p-values > 0.4; see SI-Table S9). And testing only the first judgments per face (task was repeated block wise) showed the same pattern (SI-Table S10). Moreover, the result pattern of social judgments was confirmed in a separate control study ($N = 30$) where persons were judged only once immediately after the news was encountered (see SI-page 8). We conclude that repetition did not affect the pattern of results.

2.2.2. Behavioral results in the confidence task

After the social judgment task, we conducted a separate secondary task where we investigated confidence in social judgments as a behavioral measure that reflects meta-cognitive evaluations of a given judgment. Confidence in overall social judgments showed a pattern of results like the social judgment task, replicating earlier results (Baum et al., 2020). Confidence was high for both, faces associated with negative vs. neutral and positive vs. neutral headlines independent of source credibility (Table 3 and 4, for descriptive results see SI-Table S12).

2.2.3. Event-related brain potentials in the social judgment task

EPN. For effects on relatively fast and reflexive emotional processing we investigated the EPN component (see Figure 3). We found an enhanced negativity for faces associated with negative vs. neutral headlines and that effect was not modulated by source credibility (Table 5 and 6; for all descriptive ERP results see SI-Table S13).

For faces associated with positive compared to neutral headlines the main effect and interaction with source credibility did not reach significance (Table 5), however separate tests suggest that the EPN effect was elicited for associated headlines from trusted sources, but absent for distrusted sources (Table 6).

LPP. To investigate slower, more controlled evaluative processing, we tested effects in the later LPP component (see Figure 4). For faces associated with negative headlines, we found an enhanced LPP compared to associated neutral headlines and no interaction with source credibility (Table 5). Faces associated with negative information from both, trusted and distrusted sources elicited LPP effects (Table 6).

For faces associated with positive headlines, we found an enhanced LPP compared to associated neutral headlines and no interaction with source credibility in the LPP (Table 5). Preregistered separate contrasts showed as predicted that LPP effects of faces associated with positive vs. neutral headlines were only present for trusted sources, whereas emotion effects were absent for distrusted sources (Table 6).

Post-hoc (non-preregistered), we included judgment latencies as a covariate to account for motor responses in the LPP results. This did
Phase 2: Behavioral Results
Main task: EEG Acquisition, Social Judgments

Fig. 2. Behavioral results in the social judgment task. In Phase 2 the social judgment was performed as main task to investigate the effects of source credibility and headline content. A. Participants judged persons presented in isolation based on the information from Phase 1. Faces were presented until response. B. Social judgments relied on associated emotional content for both trusted and distrusted sources, not accounting for source credibility. C. Latencies of judgments were faster than neutral and positive contents for trusted sources. For distrusted sources, only judgments based on negative headline contents were found to be faster than neutral, but not positive contents. B.-C. Raincloud plots (Allen et al., 2019) show means and 95% confidence intervals calculated with the summarySEwithin function (Morey, 2008) on single trial data, with points and distributions for data aggregated by subject.

Table 1
Mixed model summary statistics show effects of source credibility, negative and positive headline content and their interactions on behavioral dependent variables in the social judgment task. Effects on social judgments and their latencies were estimated in separate mixed models and fixed effects were coded as repeated contrasts according to our hypotheses.

| Coefficient | Social Judgments | Latencies of Judgments [−1000/latency(ms)] |
|-------------|------------------|--------------------------------------------|
|             | b (95%-CI)       | SE  | t  | p   | b (95%-CI) | SE  | t  | p   |
| Intercept (Grand Mean) | −0.03 (−0.14 − 0.09) | 0.06 | −0.45 | 0.656 | −1.37 (−1.45 − −1.29) | 0.04 | −33.06 | <0.001 |
| Source Credibility (Trusted vs. Distrusted) | −0.01 (−0.14 − 0.12) | 0.07 | −0.11 | 0.916 | 0.01 | 0.01 | 0.48 | 0.634 |
| Negative Headline Content | −1.89 (−2.13 − −1.66) | 0.12 | −16.00 | <0.001 | −0.16 | 0.03 | −4.60 | <0.001 |
| Source Credibility × Negative Headline Content | −0.03 (−0.37 − 0.31) | 0.17 | −0.16 | 0.876 | 0.01 | 0.03 | 0.41 | 0.685 |
| Positive Headline Content | 1.24 (1.04 − 1.45) | 0.11 | 11.82 | <0.001 | −0.07 | 0.03 | −2.50 | 0.016 |
| Source Credibility × Positive Headline Content | 0.12 (−0.21 − 0.45) | 0.17 | 0.71 | 0.482 | −0.01 | 0.03 | −0.31 | 0.759 |
| Model Formula | Judgment × Headline Content × Source Credibility | Latency × Headline Content × Source Credibility |
|               | Judgment × Negative Headline Content | + (S+Neg+S Neg+Pos+S' Pos | subject) + (S+Neg+S Neg+Pos+S' Pos | face) |
|               | Judgment × Positive Headline Content | + (S+Pos+S Neg+Pos+S' Pos | subject) + (S+Pos+S Neg+Pos+S' Pos | face) |

Note. Asterisk (*) stands for interaction. Abbreviations for slopes in the random effects terms: S=Source Credibility, Neg=Negative Headline Content, Pos=Positive Headline Content. Face stands for face stimulus.

Table 2
Negative and positive headline content effects on social judgments and their latencies separately within each source credibility condition computed from the models in Table 1.

| Contrast                | Social Judgments | Latencies of Judgments [−1000/latency(ms)] |
|-------------------------|------------------|--------------------------------------------|
|                         | b    | SE  | t   | p   | b    | SE  | t   | p   |
| Trusted: Neg vs. Neu    | −1.91 | 0.14 | −13.41 | <0.001 | −0.15 | 0.04 | −3.99 | <0.001 |
| Distrusted: Neg vs. Neu | −1.88 | 0.15 | −12.43 | <0.001 | −0.16 | 0.04 | −4.25 | <0.001 |
| Trusted: Pos vs. Neu    | 1.30  | 0.14 | 9.61  | <0.001 | −0.07 | 0.03 | −2.52 | 0.022 |
| Distrusted: Pos vs. Neu | 1.18  | 0.13 | 8.77  | <0.001 | −0.06 | 0.03 | −1.78 | 0.083 |
Table 3
Mixed model summary statistics show effects of source credibility, negative and positive headline content and interactions on behavioral dependent variables in the confidence task. Effects on overall judgments and on confidence ratings were estimated in separate mixed models and fixed effects were coded as repeated contrasts.

| Coefficient | Overall Judgments | Confidence in Judgments |
|-------------|-------------------|-------------------------|
|             | b (95%-CI) | SE  | t   | p   | b (95%-CI) | SE  | t   | p   |
| Intercept (Grand Mean) | –0.02(–0.14 – 0.09) | 0.06 | –0.41 | 0.688 | 69.15 (63.87 – 74.43) | 2.69 | 25.66 | <0.001 |
| Source Credibility ( Trusted vs. Distrusted) | –0.03 (–0.13 – 0.08) | 0.05 | –0.47 | 0.643 | 3.31 (–0.03 – 6.65) | 1.70 | 1.94 | 0.066 |
| Negative Headline Content ( Neg vs. Neu) | –1.91 (–2.08 – 1.74) | 0.09 | –21.92 | <0.001 | 14.07 (4.29 – 23.85) | 4.99 | 2.82 | 0.008 |
| Source Credibility * Negative Headline Content | 0.00 (–0.23 – 0.23) | 0.12 | –0.01 | 0.995 | 2.07 (–4.82 – 8.96) | 3.52 | 0.59 | 0.556 |
| Positive Headline Content (Pos vs. Neu) | 1.34 (0.115 – 1.52) | 0.09 | 14.26 | <0.001 | 17.70 (11.07 – 24.34) | 3.39 | 5.23 | <0.001 |
| Source Credibility * Positive Headline Content | 0.07 (–0.15 – 0.30) | 0.12 | 0.64 | 0.521 | –0.09 (–0.94 – 8.75) | 4.51 | –0.02 | 0.984 |
| Model Formula | Judgment ~ Headline Content * Source Credibility + (Neg + Pos || subject) + (S + Neg + Pos + S + Pos || face) | Confidence ~ Headline Content * Source Credibility + (Neg + Pos + S + Pos || subject) + (S + Neg + Pos + S + Pos || face) |

Note. Asterisk (*) stands for interaction. Double bars in random effects terms set correlation parameters to zero. Abbreviations for slopes in the random effects terms: S=Source Credibility, Neg=Negative Headline Content, Pos=Positive Headline Content. Face stands for face stimulus.

Table 4
Negative and positive headline content effects on overall judgments and confidence in judgments separately within each source credibility condition computed from the models in Table 3.

| Contrast | Overall Judgments | Confidence in Judgments |
|----------|-------------------|-------------------------|
|          | b | SE  | t   | p   | b | SE  | t   | p   |
| Trusted: Neg vs. Neu | –1.91 | 0.11 | –18.20 | <0.001 | 15.10 | 5.29 | 2.85 | 0.009 |
| Distrusted: Neg vs. Neu | –1.91 | 0.11 | –18.22 | <0.001 | 13.03 | 5.29 | 2.46 | 0.018 |
| Trusted: Pos vs. Neu | 1.38 | 0.11 | 12.45 | <0.001 | 17.66 | 4.07 | 4.34 | <0.001 |
| Distrusted: Pos vs. Neu | 1.30 | 0.11 | 11.77 | <0.001 | 17.75 | 4.07 | 4.36 | <0.001 |

Table 5
Mixed model summary statistics show effects of source credibility, negative and positive headline content and their interactions on EPN and LPP as dependent variables in the social judgment task. Effects on the predefined ROI and time range of the EPN and LPP amplitudes were estimated in separate mixed models and fixed effects were coded as repeated contrasts according to our hypotheses.

| Coefficient | EPN | LPP |
|-------------|-----|-----|
|             | b (95%-CI) | SE  | t   | p   | b (95%-CI) | SE  | t   | p   |
| Intercept (Grand Mean) | 1.16 (0.17 – 2.15) | 0.51 | 2.29 | 0.029 | 5.02 (4.28 – 5.76) | 0.38 | 13.25 | <0.001 |
| Source Credibility ( Trusted vs. Distrusted) | –0.06 (–0.26 – 0.14) | 0.10 | –0.58 | 0.569 | 0.14 (–0.05 – 0.33) | 0.10 | 1.48 | 0.149 |
| Negative Headline Content ( Neg vs. Neu) | –0.45 (–0.63 – 0.26) | 0.09 | –4.72 | <0.001 | 0.81 (0.49 – 1.14) | 0.17 | 4.88 | <0.001 |
| Source Credibility * Negative Headline Content | –0.02 (–0.43 – 0.38) | 0.21 | –0.12 | 0.907 | –0.15 (–0.55 – 0.26) | 0.21 | –0.71 | 0.479 |
| Positive Headline Content (Pos vs. Neu) | –0.17 (–0.37 – 0.03) | 0.10 | –1.69 | 0.103 | 0.37 (0.10 – 0.64) | 0.14 | 2.69 | 0.013 |
| Source Credibility * Positive Headline Content | –0.33 (–0.71 – 0.06) | 0.20 | –1.64 | 0.116 | 0.27 (–0.17 – 0.72) | 0.23 | 1.19 | 0.241 |
| Model Formula | EPN ~ Headline Content * Source Credibility + (S + S’Neg + S’Pos || subject) + (S + S’Neg + S’Pos + S’Pos || face) | LPP ~ Headline Content * Source Credibility + (S + S’Neg + S’Pos || subject) + (S + S’Neg + S’Pos + S’Pos || face) |

Note. Asterisk (*) stands for interaction. Double bars in random effects terms set correlation parameters to zero. Abbreviations for slopes in the random effects terms: S=Source Credibility, Neg=Negative Headline Content, Pos=Positive Headline Content. Face stands for face stimulus.

Table 6
Negative and positive headline content effects on EPN and LPP separately within each source credibility condition computed from the models in Table 5.

| Contrast | EPN | LPP |
|----------|-----|-----|
|          | b | SE  | t   | p   | b | SE  | t   | p   |
| Trusted: Neg vs. Neu | –0.46 | 0.14 | –3.26 | 0.006 | 0.74 | 0.20 | 3.77 | <0.001 |
| Distrusted: Neg vs. Neu | –0.43 | 0.14 | –3.09 | 0.006 | 0.89 | 0.20 | 4.52 | <0.001 |
| Trusted: Pos vs. Neu | –0.34 | 0.14 | –2.35 | 0.030 | 0.51 | 0.18 | 2.83 | 0.008 |
| Distrusted: Pos vs. Neu | –0.01 | 0.14 | –0.08 | 0.940 | 0.24 | 0.18 | 1.33 | 0.190 |
not change the effects of predictors and three-way-interactions were not significant (all $t$-values <0.7, all $p$-values >0.5; see SI-Table S14). Since all trials involved motor responses and accounting for latency differences did not affect the pattern of results, we consider systematic influences of motor response-related activity unlikely to have affected the LPP (Luck, 2014).

Additional post-hoc comparisons reveal that the positive and negative condition did not differ for trusted but for distrusted sources (trusted: $b = 0.23$, $SE = 0.22$, $t = 1.05$, $p = .299$; distrusted: $b = 0.65$, $SE = 0.22$, $t = 2.93$, $p = .009$; see SI-Table S15). A similar pattern was also found in the EPN, with the difference only for distrusted but not for trusted sources (trusted: $b = -0.12$, $SE = 0.16$, $t = -0.79$, $p = .521$; distrusted: $b = 0.42$, $SE = 0.16$, $t = -2.73$, $p = .018$; see SI-Table S15). This suggests that the valence dependent effects are directly related to the processing of the lack of source trustworthiness and not a generally smaller effect in the positive condition.

**N170.** For effects on the earlier visual processing of the faces we conducted a post-hoc analyses of the N170 (between 130 and 200 ms according to Hinojosa et al., 2015 and at the electrode sites P7, P8, PO7, PO8, PO9, PO10 according to S. Schindler et al., 2021). We found an enhanced negativity for faces associated with negative vs. neutral headlines, all other effects were not significant (Table 7, Fig. 5). Separate comparisons showed no effects (Trusted: Neg vs. Neu $b = -0.33$, $SE = 0.15$, $t = -2.19$, $p = .13$; Distrusted: Neg vs. Neu $b = -0.12$, $SE = 0.15$, $t = -0.80$, $p = .429$; Trusted: Pos vs. Neu $b = -0.20$, $SE = 0.14$, $t = -1.39$, $p = .340$; Distrusted: Pos vs. Neu $b = 0.12$, $SE = 0.14$, $t = 0.81$, $p = .429$).

### 3. Discussion

How does the brain respond to emotional information from untrustworthy sources and how does that influence our social judgments? How effective are contemplations about the credibility of media sources as a cognitive guard against the social and neural consequences of untrustworthy headlines? Here we show that simply drawing people’s attention to the trustworthiness of media sources before being exposed to news is insufficient to abolish the dominant consequences of emotional headlines regardless of how credible their source is. We asked our participants to classify media sources’ trustworthiness as a minimal cognitive intervention and guard against misinformation that can be easily applied in daily life. Subsequently, they were exposed to news headlines with social contents of neutral or emotional valence. Replicating earlier findings (Metzger and Flanagan, 2013; Pennycook and Rand, 2019; Dias et al., 2020) we observed that participants clearly discerned media sources’ trustworthiness. Despite their clear discernment of what sources they trust or distrust, when judging the persons they had read about, participants relied to a large extent on the emotional contents of the headlines irrespective of source credibility, and they did so with high confidence in their judgments. This key result is accompanied by findings indicating that this bias might be more susceptible to change for positive headline contents, whereas negative social-emotional information seems to be particularly difficult to disregard, even when we know it stems from sources we don’t trust. As further discussed below, preregistered separate tests show that effects of positive headlines were
Fig. 4. Late positive potential (LPP) modulations by headline content and source credibility effects. In Phase 2, the EEG was acquired in the social judgment task when faces were presented in isolation. We investigated neurocognitive correlates of slow and evaluative processing in the LPP (400–600 ms). **A.** The LPP was enhanced for negative contents compared to neutral for both trusted and distrusted sources, exhibiting no modulation by source credibility. **B.** For positive headline contents, the LPP was enhanced compared to neutral only for trusted sources, but not for distrusted sources. **A–B.** Grand average ERPs are shown at electrode site Pz and scalp distributions show the effects as difference between conditions in the time windows shaded in gray. **C.** Mean ERP amplitude sizes are plotted for the pre-specified regions-of-interest and time window of the LPP. Raincloud plots (Allen et al., 2019) show means and 95% confidence intervals calculated with the `summarySEwithin` function (Morey, 2008) on single trial data, and points, boxplots, and distributions for data aggregated by subject.

Table 7
Mixed model summary statistics show effects of source credibility, negative and positive headline content and their interactions on visual face processing in the N170 as dependent variable in the social judgment task.

| Coefficient                          | N170 |       |     |     |
|--------------------------------------|------|-------|-----|-----|
|                                      | b (95%-CI) | SE  | t   | p   |
| Grand Mean                           | −1.02 (−2.24 − 0.20) | 0.62 | −1.64 | 0.112 |
| Source Credibility (Trusted vs. Distrusted) | 0.11 (−0.03 − 0.26) | 0.07 | 1.55 | 0.122 |
| Negative Headline Content (Neg vs. Neu) | −0.22 (−0.42 − 0.02) | 0.10 | −2.18 | 0.037 |
| Source Credibility * Negative Headline Content | −0.21 (−0.63 − 0.22) | 0.22 | −0.96 | 0.346 |
| Positive Headline Content (Pos vs. Neu) | −0.04 (−0.26 − 0.17) | 0.11 | −0.38 | 0.706 |
| Source Credibility * Positive Headline Content | −0.32 (−0.68 − 0.05) | 0.19 | −1.70 | 0.098 |

Note. Asterisk (*) stands for interaction. Double bars in random effects terms set correlation parameters to zero. Abbreviations for slopes in the random effects terms: S=Source Credibility, Neg=Negative Headline Content, Pos=Positive Headline Content. Face stands for face stimulus.

only present for trusted but not for distrusted sources, however these findings are tentative due to the specific interactions lacking statistical significance. Our findings provide critical insights into the cognitive and brain mechanisms underlying the processing of misinformation and in the potential scope and limits of contemplating source credibility as a guard against misinformation. They will be discussed in detail below.

For positive headline contents different measures suggested signs of source credibility effects. Specifically, judgments based on positive relative to neutral contents were faster when the source was trusted, but not when the source was distrusted, possibly suggesting more hesitation when judging persons according to social-emotional information from untrustworthy sources. Furthermore, fast emotional brain responses in the EPN, a component related to more impulsive arousal-driven process-
Fig. 5. N170 modulations by headline content and source credibility effects. In Phase 2, the EEG was acquired in the social judgment task when faces were presented in isolation. We investigated neurocognitive correlates of visual face processing in the N170 (130–200 ms). A. The N170 was enhanced for negative contents compared to neutral, exhibiting no modulation by source credibility. B. For positive headline contents, no significant modulations were found. A–B. Grand average ERPs are shown at electrode site PO10 and scalp distributions show the effects as difference between conditions in the time windows shaded in gray. C. Mean ERP amplitude sizes are plotted for the pre-specified regions-of-interest and time window of the EPN. Raincloud plots (Allen et al., 2019) show means and 95% confidence intervals calculated with the summary5within function (Morey, 2008) on single trial data, and points, boxplots, and distributions for data aggregated by subject.

ing of emotional stimuli were only observed when positive (vs. neutral) headlines stemmed from trusted sources, suggesting that the early reception of positive contents depends on explicit trustworthiness evaluations of the source. Please note, however, that such an early modulation by source credibility was not predicted. Based on dual-process theories we predicted that fast brain responses were dominated by emotion and slow responses integrated both, emotional contents and source credibility. In line with these predictions, the LPP component related to later more controlled evaluative processes, sensitive to context information and deliberate control, shows effects of positive (relative to neutral) headline contents only for trustworthy sources, but not for untrustworthy sources. Thus, evaluations based on untrustworthy positive information are comparable to neutral information. Taken together, early as well as late brain responses to faces in the positive headline condition were reduced for sources judged as untrustworthy and valenced social judgments were made more slowly. Please note however that these effects of source credibility must be considered with caution, because the critical interactions of positive headline effects with source credibility lacked statistical significance, and only the planned separate tests show that effects of positive headlines were present only for trusted but not for distrusted sources. While further evidence is needed to draw clear conclusions about the susceptibility of positive headline contents to source credibility, the current results suggest that source credibility might protect against the influence of untrustworthy positive information to some extent, as effects in both fast emotional and slow, evaluative processing were restricted to trustworthy positive information.

Remarkably, fast and slow brain responses and social judgments related to negative headlines were immune to the insight into the lack of credibility before news exposure. All measures were dominated by negative contents from both trustworthy and untrustworthy sources. We found robust effects of negative relative to neutral headline contents regardless of source credibility in fast emotional responses reflected in the EPN, in later evaluative brain responses reflected in the LPP and in social judgments. Furthermore, negative headlines even affected the visual level of face processing as suggested by influences on the N170 that is related to structural encoding of and attention to faces (Giménez-Fernández et al., 2020; 2021; Krasowski et al., 2021; Luo et al., 2016; S. Schindler et al., 2021; Xu et al., 2016). These results demonstrate that negative headlines have prevailing consequences for emotional and evaluative brain responses and social judgments that withstand an explicit and deliberate contemplation of source credibility.

3.1. Limitations and future directions

What further mechanisms may play a role in the weak effects of source credibility, such as learning and memory processes related to news exposure and retrieval? One may wonder if the credibility of news was unnoticed by participants during news exposure. However, we consider this unlikely for several reasons. We show that participants can well discern the credibility of news sources. Further, each news website presentation started with showing the prominent source logo alone for one second before the face and headline information was presented, giving enough time to process the source before the headlines were presented. Additionally, we demonstrated in a control study (reported in Baum and Abdel Rahman, 2020) via tracking active eye movements that participants sample source information during the news exposure in a learning design identical to the current study. The additional explicit classifications of sources beforehand as realized here should even have increased the attention to the sources in the current study. Moreover,
the same control study showed that blurred website layouts with retracted source name are reliable cues for credibility. We are therefore confident that source credibility information was available to our participants. Like in real-life, however, it is possible that participants notice and know about source credibility when confronted with news but may not fully engage with or encode this information. This question should be targeted in future research.

After news exposure, it is possible that the source credibility information is no longer present in memory, while the emotional information is preserved. As a first step to address memory-related processes, we explored source credibility effects in dependent measures of the social judgment phase by directly considering the individual memories, that is, we assigned the source condition that the participants remembered and associated with each face in the experiment, thereby replacing the pre-defined source credibility conditions. Results of this analysis show main effects of individually remembered source credibility in several measures (Source Credibility (Trusted vs. Distrusted) in person likability after news exposure $b = 0.20$, $SE = 0.07$, $t = 2.99$, $p = .006$; social judgments $b = 0.24$, $SE = 0.09$, $t = 2.66$, $p = .011$; and confidence in judgments $b = 4.11$, $SE = 1.16$, $t = 2.56$, $p = .011$; see SI pages 17ff for results of all dependent measures). The pattern of results also confirms the differential pattern of credibility on positive and negative news headlines that our main results suggest, revealing statistically more robust effects: Additional to the differential effects of positive headlines for trusted and distrusted sources in separate comparisons (Positive Headline Content (Pos vs. Neu) with ps (0.015 for Trusted, and ps) 0.3 for Distrusted in latencies of judgments and the LPP, see SI pages 18, 20), statistical interactions of positive vs. neutral headlines with individually assigned source credibility are found in several measures (Source Credibility $\times$ Positive Headline Content was found for person likability after news exposure: $b = 0.33$, $SE = 0.16$, $t = 2.10$, $p = .036$, confidence in judgments $b = 10.96$, $SE = 4.57$, $t = 2.40$, $p = .024$, and the LPP $b = 0.68$, $SE = 0.30$, $t = 2.27$, $p = .029$; see SI pages 17f). Whereas the influence of associated negative headlines was not modulated by source credibility assignments (Source Credibility $\times$ Negative Headline Content all ps > 0.3, see SI pages 17f).

These additional findings suggest that memory plays a role especially for the effects of source credibility. However, the general pattern of the findings matches the main results, supporting a differential effect for positive and negative headline contents. This corroborates the conclusion that the bias to disregard source credibility may be less pronounced in the positive headline condition, suggesting that it is the dominance especially of the negative information that weakens the influence of source credibility. However, future research is needed, especially to investigate the relative contribution of learning and memory mechanisms more directly.

The dominant consequences of emotional contents of news despite its credibility is likely a general mechanism underlying the current findings and unlikely due to the specific paradigm of the study. We employed a controlled experimental design to systematically manipulate news credibility and content without confounding factors while enhancing ecological validity by using well-known media sources and original layouts that closely resemble news media encountered in real life. As discussed earlier, we can assume that the news’ credibility was noticed by participants during news exposure. Additionally, we can show that the experimentally necessary quantity of faces and repetitions in the judgment task are unlikely to have caused our results, for example by participants sorting persons into broad categories. We can show the same pattern of dominant effects of emotional information in a control experiment with immediate, one-time judgments after each news was presented, thereby removing repetitions in both, exposure, and judgments (see Results section and SI page 8). Moreover, accounting for repetitions of the judgment task in the analysis of our current results revealed the same findings. Furthermore, the tendency for differential effects of source credibility for positive contents but not negative contents speaks against a general explanation due to the experimental procedure or strategic responses on the side of the participants. Therefore, we conclude that a general bias due to the experimental design is unlikely and that, on the flipside, the controlled setup strengthens the robust neurocognitive evidence that we aimed for.

We examined the neural and social consequences of news on relatively global social evaluations that reflect a natural tendency of forming social impressions from visual appearance or semantic information (Todorov et al., 2007; Bliss-Moreau et al., 2008; Uhlmann et al., 2015). It is possible that participants focused more on the emotional information than the source of news to make these judgments. We instructed participants to base their judgments on all information of Phase 1. This implicitly includes news credibility but does not explicitly mention it. Like in real life when confronted with emotional information about others, it is left to us to consider the available trustworthiness, but we might not necessarily use this information to put our judgments into perspective. One may ask if the specific task or instruction led to dominant effects of emotional contents, and if other tasks or instructions result in other outcomes. Here, we address this concern by additionally asking participants for their confidence in judgments and results show again a reliance on emotional contents, suggesting a general mechanism. Future research must show if making source credibility explicit in the task or instruction helps to reduce the dominant effects of emotional contents, for example by instructing participants to base their judgments on the available contents of news and the credibility of the media sources. However, we think that such tasks create situations that differ from how person judgments are made in natural situations as described above. Moreover, we used high-temporally resolving ERPs to investigate the different levels of information processing in response to news. Additional EEG analyses and other neuroimaging measures may show different and complementary results. For example, analyzing social relevance via induced activity in frequency bands, investigating signals in different neural systems via functional brain imaging, and employing connectivity measures would add valuable evidence to the picture (e.g., Ensenberg et al., 2017, Gordon et al., 2019; Van Bavel et al., 2015).

Our findings provide experimental insight into the precise neurocognitive mechanisms that may underlie real life behaviors observed in online media, enhancing our understanding of how misinformation is processed at an individual level and extending recent research. Fast and slow brain responses revealed the mechanisms underlying the “success” of emotional headlines irrespective of their trustworthiness. Thereby, our work suggests that emotions in headlines bias the information processing system even against better knowledge. This partially contradicts predictions we derived from classical dual-process theories assuming that only fast processing, but not slow and more controlled processing is subject to such influences. Further, the present findings advance certain aspects of the concept of nudging interventions in the context of (mis)information (e.g., Kozyrева et al., 2020; Lorenz-Spreen et al., 2020; Pennycook et al., 2020; Pennycook et al., 2021; Roosenbeek and van der Linden, 2019) by suggesting that nudges targeting the contemplation of news trustworthiness are limited in warding off the influences of emotional contents. Specifically, other nudging interventions asked participants to think about the accuracy of one headline on a general topic and found that participants improved in judging the accuracy of headlines related to political and health information (Pennycook et al., 2021, 2020), whereby source names were also presented along with the verbatim headlines from published articles (e.g., New York Times, Washington Post, Fox News, or Washington Examiner). Like the current intervention this nudge was aimed at increasing attention to the trustworthiness of news. However, important differences between the approaches pertain thinking about the trustworthiness in relation to a headline vs. media sources; and the outcome measures, that is, judging the accuracy of the news vs. making judgments based on what the news alleges. These differences and similarities highlight interesting paths for future investigations. Finally, understanding these mechanisms can be viewed as a first step to develop further cognitive strategies as protection against misinformation, which is relevant for testable hypotheses in various dis-
Acknowledgements

We thank Guido Kiecker for technical support and Friedrich Eisenbeck, Alexander Enke, Markus Fiukowski, Magdalena Kleissner, Jan-Luca Neumark, Wiebke Rehn, Kirsten Stark, Martin Zielinski, and Anja Zierep for assisting in stimulus preparation and/or data acquisition. We thank our reviewers for valuable reviews.

Funding

This work was supported by a PhD scholarship of Studienstiftung des Deutschen Volkes to Julia Baum and a German Research Foundation grant AB 277–6 to Rasha Abdel Rahman.

Data Availability

Data and code is available online (osf.io/hqwy2).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2021.118572.

References

Abdel Rahman, R., 2011. Facing good and evil: early brain signatures of affective biographical knowledge in face recognition. Emotion 11 (5), 1397–1405. doi:10.1037/a0024717.

Allen, M., Poggiagli, D., Whitaker, K., Marshall, T.R., Kievet, R.A., 2019. Raincloud plots: a multi-platform tool for robust data visualization. Wellcome Open Res. 4, 63. doi:10.12688/wellcomeopenres.15191.1.

Anderson, E., Siegel, E.H., Bliss-Moreau, E., Barrett, L.F., 2011. The visual impact of gossip. Science 332 (6026), 1446–1449.

Aristei, S., Knoop, C.A., Lubrich, O., Nehrling, T., Enge, A., Stark, K., Sommer, W., & Abdel Rahman, R. (2021). Affect as Anaesthetic. How emotional contexts modulate the processing of counterintuitive concepts. [Manuscript submitted for publication]

Aviezer, H., Eisenberg, N., Hassin, R.R., 2017. The inherently contextual nature of facial emotion perception. Curr Opin Psychol 17, 47–54. doi:10.1016/j.copsyc.2017.06.006.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, R.H. (2015a). Parsimonious mixed models. Available from arXiv:1506.04967 (stat.ME).

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015b. Fitting linear mixed-effects models using lme4. J. Statist. Softw. 67 (1), 1–48. doi:10.18637/jss.v067.i01.

Baum, J., Abdel Rahman, R., 2020. Emotional news affects social judgments independent of perceived media credibility. Soc. Cogn. Affect. Neurosci. 11 (6), 1397. doi:10.1093/scan/snaa164.

Baum, J., Frömer, R., Abdel Rahman, R., 2020. Emotions and cognitive effort - but not source credibility - determine news-based social judgments. PsyArxiv doi:10.31234/osf.io/2x294.

Baum, J., Rabovsky, M., Rose, S.B., Abdel Rahman, R., 2018. Clear judgments based on unclear evidence: public opinion is strongly influenced by untrustworthy gossip. Emotion 20 (2), 246–260. doi:10.1037/emo0000545.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Statist. Soc. Series B (Methodological) 57 (1), 289–300. 10.2307/2346107; refsearch-gateway: 025a191901052c998b35856aa9a16565d.

Berg, P., Scherg, M., 1991. Dipole models of eye movements and blinks. Electroencephalography Clin. Neurophysiol. 79 (1), 38–44. doi:10.1016/0164-6644(91)90154-V.

Blechert, J., Shepper, G., Di Tella, C., Williams, H., Gross, J.J., 2022. See what you think: reappraisal modulates behavioral and neural responses to social stimuli. Psychol. Sci. 33 (4), 346–353. doi:10.1177/0956797621438559.

Bliss-Moreau, E., Barrett, L.F., Wright, C.I., 2008. Individual differences in learning the affective value of others under minimal conditions. Emotion 8 (4), 479–493. doi:10.1037/1528-3542.8.4.479.

Brady, W.J., Gantman, A.P., Van Bavel, J.J., 2020. Attentional capture helps explain why moral and emotional content go viral. J. Experiment. Psychol. 149 (4), 746–756. doi:10.1037/xge0000673.

Brady, W.J., Wills, J.A., Jost, J.T., Tucker, J.A., Van Bavel, J.J., 2017. Emotion shapes the diffusion of moralized content in social networks. Proc. Natl. Acad. Sci. 114 (28), 7313–7318. doi:10.1073/pnas.1618923114.

Crockett, M.J., 2017. Moral outrage in the digital age. Nature Human Behav. 1 (3). doi:10.1038/s41562-017-0213-3.

Cunningham, W.A., Zelazo, P.D., 2007. Attitudes and evaluations: a social cognitive neuroscience perspective. Trends Cogn. Sci. (Regul. Ed.) 11 (3), 97–104. doi:10.1016/j.tics.2006.01.009.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134 (1), 9–21.
Skowronski, J.J., Carlton, D.E., 1987. Social Judgment and Social Memory: the Role of Cue Diagnosticity in Negativity, Positivity, and Extremity Biases. J. Pers. Soc. Psychol. 52 (4), 689–699. doi:10.1037/0022-3514.52.4.689.

Soroka, S., Fournier, P., Nir, L., 2019. Cross-national evidence of a negativity bias in psychophysiological reactions to news. Proc. Natl. Acad. Sci. 116 (38), 18888–18892.

Stein, T., Grubb, C., Bertrand, M., Suh, S.M., Verosky, S.C., 2017. No impact of affective person knowledge on visual awareness: Evidence from binocular rivalry and continuous flash suppression. Emotion 17 (8), 1199–1207. doi:10.1037/emo0000305.

Strack, F., Deutsch, R., 2004. Reflective and Impulsive Determinants of Social Behavior. Pers. Soc. Psychol. Rev. 8 (3), 220–247. doi:10.1207/s15327957pspr0803_1.

Suess, F., Rabovsky, M., Abdel Rahman, R., 2014. Perceiving emotions in neutral faces: expression processing is biased by affective person knowledge. Soc. Cogn. Affect. Neurosci. 10 (4), 531–536. doi:10.1093/scan/nsu088.

Todorov, A., Gobbin, I.I., Evans, K.K., Haxby, J.V., 2007. Spontaneous retrieval of affective person knowledge in face perception. Neuropsychologia 45 (1), 163–173. doi:10.1016/j.neuropsychologia.2006.04.018.

Todorov, A., Olson, L.R., 2008. Robust learning of affective trait associations with faces when the hippocampus is damaged, but not when the amygdala and temporal pole are damaged. Soc. Cogn. Affect. Neurosci. 3 (3), 195–203. doi:10.1093/scan/nsn013.

Uhlmann, E.L., Pizarro, D.A., Diermeier, D., 2015. A Person-Centred Approach to Moral Judgment. Perspect. Psychol. Sci. 10 (1), 72–81. doi:10.1177/1745691614556679.

Van Bavel, J.J., Baicker, K., Boggio, P.S., Capraro, V., Cichocka, A., Cikara, M., Crockett, M.J., Crum, A.J., Douglas, K.M., Druckman, J.N., Drury, J., Dube, O., Ellemers, N., Finkel, E.J., Fowler, J.H., Gelfand, M., Han, S., Haslam, S.A., Jetten, J., ... Willer, R., 2020. Using social and behavioural science to support COVID-19 pandemic response. Nature Human Behav. 1–12.

Van Bavel, J.J., FeldmanHall, O., Mende-Siedlecki, P., 2015. The neuroscience of moral cognition: from dual processes to dynamic systems. Curr. Opin. Psychol. 6, 167–172. doi:10.1016/j.copsyc.2015.08.009.

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

Wieser, M.J., Gerdes, A.B.M., Büngel, I., Schwarz, K.A., Mühlberger, A., Pauli, P., 2014. Not so harmless anymore: how context impacts the perception and electrocortical processing of neutral faces. Neuroimage 92 (C), 74–82. doi:10.1016/j.neuroimage.2014.01.022.

Xu, M., Li, Z., Diao, L., Fan, L., Yang, D., 2016. Contextual valence and sociality jointly influence the early and later stages of neutral face processing. Front. Psychol. 07 (1368), 1446. doi:10.3389/fpsyg.2016.01368.

Yonelinas, A.P., 2002. The nature of recollection and familiarity: a review of 30 years of research. J. Mem. Lang. 46 (3), 441–517. doi:10.1006/jmla.2002.2864.