Emotional content in social misinformation affects mind, brain, and judgments

Dissertation

zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)
im Fach Psychologie

eingereicht an der Lebenswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von
Julia Baum, M.Sc.

Präsident der Humboldt-Universiät zu Berlin (kommissarisch): Prof. Dr. Peter Frensch
Dekan der Lebenswissenschaftlichen Fakultät: Prof. Dr. Dr. Christian Ulrichs

Gutachter*innen:

1. Prof. Dr. Rasha Abdel Rahman
2. Dr. Sebastian Schindler
3. Prof. Dr. Reinhard Beyer

Datum der Verteidigung: 11. August 2022

Contents

Abstract 3
Zusammenfassung 4
1. Introduction 5
1.1 A theoretical framework of fast and slow information processing 6
1.2 A neurocognitive perspective on the effects of misinformation 7
1.3 How social information affects person perception and social judgments 9
1.4 Prevalent ways to communicate and assess the credibility of news 11
1.5 The experimental paradigm of news exposure and derived social judgements 12
1.6 Hypotheses for the effects of emotional content and credibility 13
2. Summary of the present studies 15
2.1 Study 1: Clear judgments based on unclear evidence: Person evaluation is strongly influenced by untrustworthy gossip 15
2.2 Study 2: Emotional news affects social judgments independent of perceived media credibility 16
2.3 Study 3: Negative news dominates fast and slow brain responses and social judgments even after source credibility evaluation 17
3. General Discussion 19
3.1 A general mechanism of the dominant influence of emotional content regardless of credibility 20
3.1.1 Emotional content drives the fast reception of news 20
3.1.2 Emotional content, not credibility, dominates the slow evaluation of relevance and meaning 20
3.1.2 Credibility effects might depend on the valence of contents 21
3.1.4 What social misinformation means for person perception and judgments 22
3.2 Experimental neurocognitive evidence underlying real-life behaviour 23
3.3 On advancing the understanding of the susceptibility to misinformation 24
3.4 Limitations and future research 26
3.5 Final remarks 28
References 29
Danksagung 40
Eigenständigkeitserklärung 42
Original Articles 43

Abstract

How does misinformation influence our mind, brain, and judgments? Misinformation, also called “fake news”, is highly prevalent in massive online communication affecting public discourse and social coexistence. However, little is known about how the human information processing system is affected by misinformation on the individual level when we derive opinions and judgments. This dissertation investigates the cognitive and brain mechanisms underlying the processing of social, person-related misinformation. Within a conceptual framework of fast and slow information processing, we investigate the precise role of the emotional content and the credibility of news using high temporally resolving electrophysiological brain responses. Three studies employed real-life indicators of news credibility. Participants were exposed to negative, positive, or relatively neutral news about other persons that was either verbally marked as untrustworthy by adding e.g., “allegedly” (Study 1), or stemmed from well-known media sources perceived as trusted or distrusted (Study 2 and 3). Subsequently, we found that social judgments of the persons from the news strongly relied on the emotional content independent of the credibility, showing how social misinformation affects person evaluation although it is perceived as untrustworthy. Electrophysiological indexes of early emotional and arousal-related processes, as well as correlates of later evaluative processing were enhanced for persons associated with emotional contents regardless of the credibility of the information. This shows the pronounced influence of emotional contents not only on the initial and early response to news, but even on processes that were expected to evaluate the information on merit of its credibility. As a general mechanism these findings reveal that people’s social judgments and their fast-impulsive as well as their slow-controlled information processing are dominated by the emotional contents of news irrespective of credibility. In a first attempt to intervene, participants were instructed to explicitly evaluate the credibility of the source before reading the headlines (Study 3). This helped to overcome the bias for positive news and process its credibility to some degree. However, remarkably the insight into the lack of credibility had no influence on the effects of negative news contents on brain responses and social judgments. Our results demonstrate how emotional content in social misinformation can affect mind, brain, and judgments even against better knowledge of its lacking credibility. This also explains why “fake news” is so “successful”. In perspective, these insights help to face the challenges of misinformation from the perspective of the individual’s cognition.

Zusammenfassung

Wie beeinflusst Mis- und Desinformation unseren Verstand, unser Gehirn und unser Urteilsvermögen? Misinformation, auch „Fake News“ genannt, ist in der massiven Online-Kommunikation weit verbreitet und beeinflussen den öffentlichen Diskurs und das gesellschaftliche Zusammenleben. Es ist jedoch wenig darüber bekannt, wie Misinformation das menschliche Informationsverarbeitungssystem auf individueller Ebene beeinflusst, wenn wir Meinungen bilden und Urteile ableiten. In dieser Dissertation werden die kognitiven und Gehirn Mechanismen untersucht, die der Verarbeitung sozialer, personenbezogener Misinformation zugrunde liegen. Innerhalb eines konzeptionellen Rahmens der schnellen und langsamen Informationsverarbeitung untersuchen wir die genaue Rolle des emotionalen Gehalts und der Glaubwürdigkeit von Nachrichten anhand von zeitlich hochauflösenden elektrophysiologischen Gehirnreaktionen. In drei Studien wurden Indikatoren für die Glaubwürdigkeit von Nachrichten aus dem täglichen Leben verwendet. Proband*innen wurden mit negativen, positiven oder relativ neutralen Nachrichten über andere Personen konfrontiert, die entweder verbal als unglaubwürdig gekennzeichnet waren, z. B. durch den Zusatz "angeblich" (Studie 1), oder aus bekannten Medienquellen stammten, die als glaubwürdig oder unglaubwürdig wahrgenommen wurden (Studie 2 und 3). Daraufhin fanden wir heraus, dass soziale Urteile über die Personen stark vom emotionalen Gehalt der Nachrichten beeinflusst waren, unabhängig von der Glaubwürdigkeit. Das zeigt, wie soziale Misinformation die Bewertung von Personen beeinflusst, obwohl sie als nicht vertrauenswürdig wahrgenommen wird. Elektrophysiologische Korrelate früher emotionaler und erregungsbezogener Prozesse sowie Korrelate späterer evaluierender Verarbeitung waren verstärkte für Personen, die mit emotionalen Inhalten assoziiert wurden—unabhängig von der Glaubwürdigkeit der Information. Dies zeigt den ausgeprägten Einfluss emotionaler Inhalte nicht nur auf die unwillkürliche und frühe Reaktion auf Nachrichten, sondern sogar auf Prozesse, für die erwartet wurde, dass sie die Information aufgrund ihrer Glaubwürdigkeit evaluieren würden. Als allgemeinen Mechanismus zeigen diese Befunde, dass unsere sozialen Urteile und unsere schnell-impulsiven sowie langsam-kontrollierten Gehirnreaktionen von den emotionalen Inhalten der Nachrichten dominiert werden, unabhängig von der Glaubwürdigkeit. In einem ersten Versuch zu intervenieren, wurden die Teilnehmer*innen instruiert, vor der Konfrontation mit Schlagzeilen die Glaubwürdigkeit der Quelle explizit zu bewerten (Studie 3). Dies trug zu einem gewissen Grad dazu bei, den Bias für positive Nachrichteninhalte zu überwinden und deren Glaubwürdigkeit zu verarbeiten. Bemerkenswerterweise hatte die Einsicht in die fehlende Glaubwürdigkeit jedoch keinen Einfluss auf die Effekte negativer Nachrichteninhalte auf Gehirnreaktionen und soziale Urteile. Unsere Ergebnisse zeigen, wie der emotionale Gehalt sozialer Misinformation den Verstand, das Gehirn und das Urteilsvermögen beeinflussen kann, selbst wider besseres Wissen über die fehlende Glaubwürdigkeit. Das erklärt auch, warum „Fake News“ so „erfolgreich“ sind. Perspektivisch helfen diese Erkenntnisse, uns den Herausforderungen von Mis- und Desinformation aus Sicht der individuellen Kognition zu stellen.

1. Introduction

You may have experienced that a news headline you read or a rumour you heard about someone left you feeling outraged towards them and their behaviour. The credibility and veracity of second-hand information can however vary greatly. What you learn about others may be true and valuable—or its meaning may be dubious and without value. Here, misinformation is defined as such information that lacks credibility and may take shape as baseless gossip or headlines from untrustworthy sources perpetuating misleading or incorrect half-truths up to outright lies. How does such information affect your mind, brain, and judgments? Although stories about other people is a major subject in human communication (Dunbar, 2004; Dunbar et al., 1997; Robbins & Karan, 2020), little is known about how we are affected by social misinformation (Fig. 1 illustrates a real-world example of social misinformation).

A growing research field is striving to catch up with the phenomenon of misinformation, as misinformation and "fake news" poses an increasing problem in today's massive online communication. Once shared, unintentional or intentional misinformation can spread like wildfire and reach a global audience, shaping people's opinions and public discourse (Lazer et al., 2018; Lewandowsky et al., 2012; Vosoughi et al., 2018). By mainly investigating large-scale observational and behavioural data, research has largely focused on how information cascades through social networks, how well crowds and people can discern true from false news, what information they share, and what interventions or algorithms could contain the spread of misinformation and correct it (Bhadani et al., 2022; Ecker et al., 2022; Kozyreva et al., 2020; Lorenz-Spreen et al., 2020; Pennycook & Rand, 2021; Vosoughi et al., 2018). However, despite the prevalence of misinformation and its potentially detrimental effects, much remains unknown about the impact of misinformation on an individual's mind and brain.

This dissertation investigates the neuro-cognitive underpinnings of how the individual's cognition and judgments are affected by information holding or lacking credibility. Thus, it adds a new and complementary perspective to the research on misinformation. Electrophysiological signatures of brain activity can reveal the consequences of (mis)information on the microstructure of information processing. By considering these consequences within a framework of fast and slow information processing, this work elucidates what cognitive processes are influenced how by which aspects of (mis)information at what point in the processing stream. Understanding the cognitive processes that lead to judgments based on misinformation advances a comprehensive picture of our susceptibility to it, why it prevails in our communication, and what we can do to protect ourselves against its influence.

Across three studies, this work focuses on the precise role of emotional content. Despite possibly being a key factor to the "success" of misinformation, emotional content has received little attention in research. The way we process emotional content and its credibility may in fact help explain large-scale observations, such as that false news with fear-evoking contents travel faster and further in social networks than true news that is more characterized by contents evoking anticipation or joy (Vosoughi et al., 2018).

Figure 1. This inflammatory headline published by BILD on the 4th of December 2021 is a real-life example of person-related, social misinformation during the Corona pandemic. The headline claims that the depicted scientists are responsible for putting people in lockdown, and for bringing frustration and misery upon them. Contrary to the assumption that such headlines do little harm as "no one trusts that source anyways", we show how it can affect people's opinions and judgments, nevertheless. Universities hence have condemned this vilification of scientists in the media[1].

1.1 A theoretical framework of fast and slow information processing

The theoretical framework of this work conceptualizes how the emotional contents and the credibility of news may be processed along an information processing stream in different cognitive processes (see Fig. 2 for a conceptual overview). The core concept is based on the family of dual-process theories that generally differentiate between faster and slower systems or interactive processes (Gawronski & Bodenhausen, 2006; Kahneman, 2003; Lieberman, 2007; Strack & Deutsch, 2004). Concerning fast systems or processes, theoretical models assume so-called associative, reflexive, and relatively automatic processing of information (e.g., Gawronski & Bodenhausen, 2006; Strack & Deutsch, 2004). As such, fast processing is mostly insensitive to broader semantic relations that refine the meaning of information. For example, news about a person's vile behaviour would put them reflexively in a bad light. In comparison, concerning slow systems or processes, these models assume so-called propositional, reflective, and controlled processing of information. Notably, slow processing is sensitive to the broader and elaborate meaning of the information including its validity and truth-value (De Houwer et al., 2020a). For example, your lack of trust in the news source would theoretically devalue the meaning of the outrageous information in the headline and move the person into a more neutral light.

With respect to the processing of *emotional* information, appraisal theories of emotion assume processing levels that can be considered in analogy with the concept of fast and slow processing (Ellsworth & Scherer, 2003; Scherer, 2001). At initial and rapid processing levels, stimuli are checked for a coarse detection of emotional salience, focused on the intrinsic pleasantness and arousal of the stimuli. This is followed by more controlled and elaborate assessments regarding the relevance and implications for the observer's well-being, the coping possibilities that arise, and evaluations of the normative significance, like the agreement with moral values or the validation of beliefs.

One aspect that shapes how emotional information is assessed is its positive or negative valence. For instance, it is possible that threatening negative information is evaluated as relevant even when the implications for the observer could be reappraised as

[1] https://www.hu-berlin.de/en/press-portal/nachrichten-en/december-2021/nr-21128?set_language=en

minor, for example due to its context (Cacioppo et al., 1999; Kahneman & Tversky, 1979). This might happen reflexively or deliberately and as a protection against the potential threat a stimulus might pose (e.g., Öhman & Mineka, 2001).

Theories regarding the processing of *social* information also embody the concept of fast and slow processes e.g., (Cunningham & Zelazo, 2007; Ehret et al., 2015; FeldmanHall & Shenhav, 2019; Lieberman, 2007; Lieberman et al., 2002; Strack & Deutsch, 2004). When we evaluate other persons, initial processes or processing iterations are assumed to be relatively automatic, while additional processes or iterations and more control-demanding processes incorporate more information over time, reaching a more elaborate person evaluation or inference. Contemporary models of face perception and recognition suggest that emotional information or knowledge about persons can affect face perception already at early stages. Such parallel models assume that perceptual and cognitive processing interact based on the neuronal architecture of the brain (Gobbini & Haxby, 2007; Haxby et al., 2000). Via mechanisms of predictive processing social knowledge such as associated news influences early and later processing of faces (e.g., Otten et al., 2017). This includes affective predictions generated from past visceral experiences that can shape already the visual perception of faces (Abdel Rahman, 2011; Barrett & Bar, 2009). This is in contrast to strictly serial models assuming that information associated with a person can only be accessed at later processing levels (Bruce & Young, 1986).

Further cognitive capacities likely play a role for news-based social judgments, such as memory-related processes. In analogy to the core concept, models specific to recognition and memory retrieval similarly distinguish faster and slower retrieval, where slower processes retrieve additional context and source information that may be stored unitized or separately (Yonelinas, 2002).

Based on this general theoretical framework of fast and slow information processing, we hypothesize differential effects of the emotional contents and the credibility of news (Fig. 2). Fast processing may primarily be affected by the emotional salience and arousal of news, whereas slow processing should be sensitive to the meaning of emotion in a given context and should therefore be sensitive to the credibility of the news. The credibility of news is assumed to determine the relevance of the emotional information to the observer, their values, and their judgment. Thus, the emotional contents in untrustworthy news should affect us less than those in trustworthy news. Ultimately, the emotional contents of untrustworthy news should count little for accurate and fair explicit judgments of persons and their social character.

1.2 A neurocognitive perspective on the effects of misinformation

The neuro-cognitive underpinnings of the effects of misinformation are considered within the theoretical framework of fast and slow information processing (see Fig. 2). High-temporally resolved event-related brain potentials (ERPs) of the electroencephalogram (EEG) allow us to track how the processing of different aspects of news unfolds over time. To pinpoint what aspects of news affect which processes, we use an experimental approach that offers full control of the manipulation of the emotional

content and the credibility of news (see section 1.4). Building on accumulated past evidence that has described the functional significance of specific ERP components, we investigate distinct earlier and later cognitive processes.

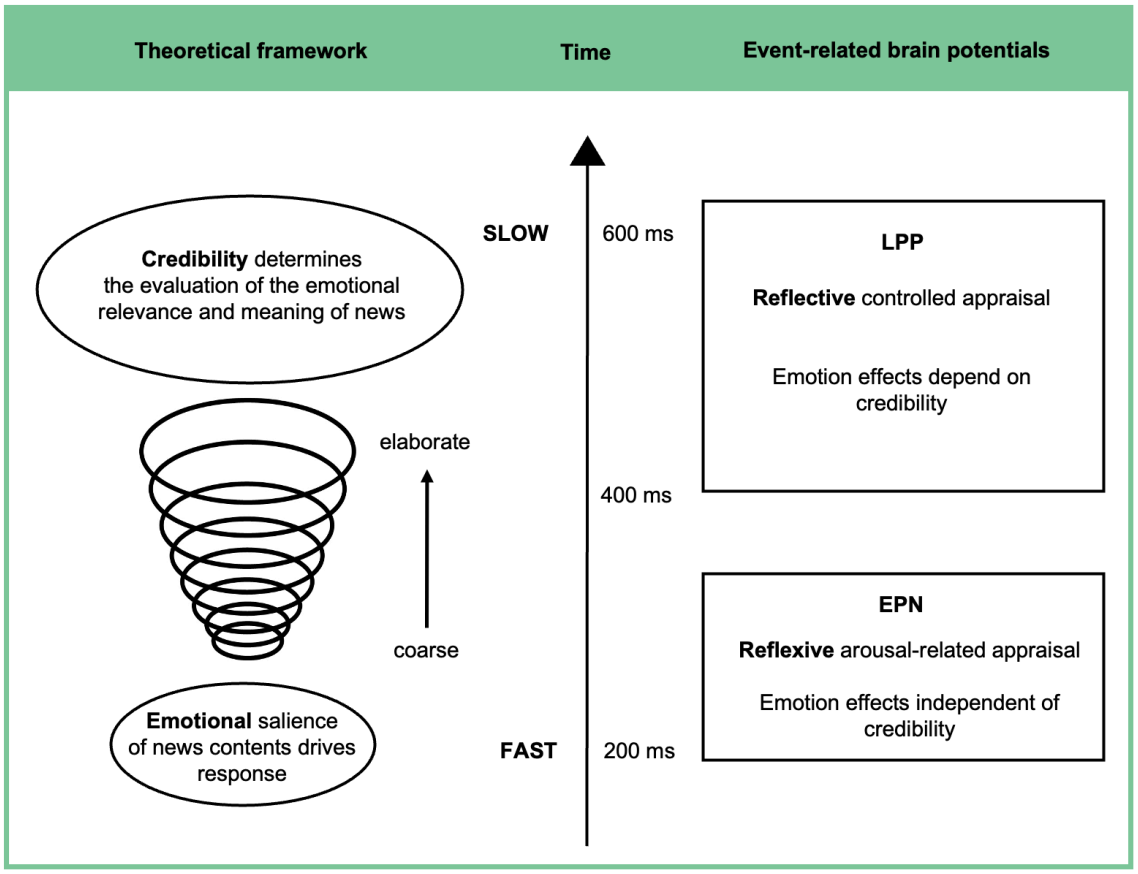

Figure 2. The processing of (mis)information: A conceptual overview of the theoretical framework of the information processing stream with central event-related potential (ERP) components, and the respective hypotheses. In the information processing stream, fast processes respond to the coarse emotional salience and arousal of news. Slow processes elaborate the meaning and relevance of news in perspective of its credibility. Information processing along the time axis is reflected in temporally high resolving ERPs. Fast processing in the EPN shows reflexive appraisal that is expected to show emotion effects independent of credibility. Slow processing in the LPP shows reflective appraisal that is expected to reflect emotion effects modulated by credibility, with reduced or absent emotion effects for untrustworthy news.

The early posterior negativity (EPN) and the late positive potential (LPP) are two ERP components whose temporal and functional attributes are well suited to address the current key questions concerning the fast and slow processing of the emotional contents and the credibility of news. 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. Sabatinelli et al., 2013; Schacht & Sommer, 2009b; Schupp et al., 2004). The EPN has been found to be primarily affected by emotional

salience and arousal relatively independent of task demands and the relevance of emotional contents in context; whereas effects on the LPP were shown to be sensitive to additional information that puts the meaning or relevance of emotion in perspective of a given context, such as a current task, goal, or revised appraisal (Blechert et al., 2012; Hajcak et al., 2010; Herbert et al., 2013; Herbert, Pauli, et al., 2011; Rellecke et al., 2012; Schacht & Sommer, 2009a; Schindler et al., 2019, 2020, 2021; Schindler & Bublatzky, 2020). Accordingly, emotion effects of the contents of news should be reflected in the EPN still independent of credibility. In comparison, emotion effects in the LPP should depend on the credibility of news, resulting in reduced or absent emotion effects for untrustworthy news (please see section 1.5 for more detailed hypotheses).

In sum, the EPN and LPP component can shed light on cognitive processes that are dissociable due to their timing and functional significance. Next, previous evidence has shown that both the EPN and LPP can be modulated by verbal social-emotional information in response to the face associated with it. This is relevant for how social-emotional *mis*information may be processed. The following section discusses what we know so far about how associated semantic information is processed and how that could be affected by the credibility of the information.

1.3 How social information affects person perception and social judgments

We effortlessly and routinely form impressions and opinions about others from second-hand semantic information (Abdel Rahman, 2011; Bliss-Moreau et al., 2008; Suess et al., 2013; Uhlmann et al., 2015). Sparse declarative, verbally transmitted person-related information is enough to learn the affective value of others via mechanisms of verbal evaluative learning and person attribution (for example, “He bullied his apprentice”; Bliss-Moreau et al., 2008; Mattarozzi et al., 2015; Todorov & Olson, 2008; Xu et al., 2016). This appears to apply to information specifically attributed to a person, and not generally merely affective information (Ferrari et al., 2020). Neurocognitive evidence suggests that during encoding and evaluative learning about depicted persons, face perception is affected at earlier and later processing stages (Junghöfer et al., 2017; Kissler & Strehlow, 2017). For affective word encoding per se, the relevance of emotion words for a given verbal context is processed at later, more elaborate stages, whereas early, more automatic emotion processing may initially be relatively independent of such contexts (Fields & Kuperberg, 2012; Herbert, Herbert, et al., 2011; Herbert, Pauli, et al., 2011; Rohr & Abdel Rahman, 2018). For example, Rohr & Abdel Rahman (2018) found that when emotional words described persons and therefore were socially relevant, later evaluative processing in the LPP was enhanced compared to non-socially relevant emotional words. However, there is also evidence that taking emotional words out of isolation and into meaningful communicative social context can enhance emotion effects already in earlier ERPs (Rohr & Abdel Rahman, 2015; Schindler et al., 2015; Schindler & Kissler, 2016).

Faces that have been associated with verbal social-emotional news have gained intrinsic emotional relevance that influences how we process them. Neurocognitive evidence shows that associated social-emotional information can affect both early and

later stages of face processing as proposed by parallel models of face perception, thus shaping how faces are perceived and persons are evaluated (e.g., Gobbini & Haxby, 2007; Otten et al., 2017; see also Schindler & Bublatzky, 2020; Schweinberger & Neumann, 2016). This is reflected in early neural signatures of enhanced reflexive and arousal-related face processing (EPN), as well as slow, later correlates of elaborate person evaluation (LPP); it influences how we see and interpret facial expressions, how likable we find others, and how we judge their character (Abdel Rahman, 2011; Aviezer et al., 2017; Galli et al., 2006; Goodwin et al., 2014; Junghöfer et al., 2017; Kissler & Strehlow, 2017; Klein et al., 2015; Luo et al., 2016; Schindler et al., 2021; Suess et al., 2013; Wieser et al., 2014; Wieser & Brosch, 2012; Xu et al., 2016). Also here, it was shown that specifically socially relevant affective information modulates the EPN and LPP, but not generally merely affective information that is non-social, e.g., "cut his finger" (Xu et al., 2016). These findings imply, for example, that facial expressions are seen as rather negative when the person is associated with negative biographical information, such as them violently persecuting minorities, although their face displays neutral expressions according to their facial muscle configuration and ratings of their expressions when the observer is unaware of their actions (Abdel Rahman, 2011; Suess et al., 2014). Electrophysiological evidence suggests that these modulations due to emotional person knowledge happen reflexively and can affect the visual perception of faces in a top-down manner (Abdel Rahman, 2011; Suess et al., 2014). Indeed, faces associated with negative social information and other prejudice have been found to elicit effects on even early visual processing in the N170 component that is related to structural encoding of and attention to faces (Giménez-Fernández et al., 2020, 2021; Hinojosa et al., 2015; Krasowski et al., 2021; Luo et al., 2016; Schindler et al., 2021; Xu et al., 2016). Emotional person-related information even increases the chances that a face is consciously seen at all by lending it privileged access to visual consciousness at early processing levels (Anderson et al., 2011; Eiserbeck et al., 2021; Eiserbeck & Abdel Rahman, 2020; but see for conflicting findings Rabovsky et al., 2016; Stein et al., 2017).

Effects of affective person-related information have been found for both positive and negative contents. However, negative contents might be processed under higher priority and be considered higher in social informational value and more diagnostic for a person's social character, as previously seen in stronger effects of negative compared to positive information (Abdel Rahman, 2011; Anderson et al., 2011; Schupp et al., 2004; Suess et al., 2013; Suzuki et al., 2013).

In sum, considering the rich effects of social-emotional information on early and later cognitive processing and behaviour, social-emotional misinformation can potentially have detrimental consequences. Some evidence suggests that these effects may depend on how meaningful the information is in a given context, for example if it is social or non-social. However, it is unclear how meaningful person-related information is processed when it is of questionable veracity, and surprisingly little is known about how the credibility modulates the effects of person-related information. We tackle this question by investigating the effects of the credibility of information as it is communicated and assessed in everyday life.

1.4 Prevalent ways to communicate and assess the credibility of news

We investigate two prominent and widely used indicators of how trustworthy and credible news is: Verbal marking and knowledge about the credibility of media sources. While both are common, they have different properties and functions. It is important to study these different ways to identify the credibility of news, because they may differentially affect processing and judgments, or we may find a pattern that suggests a general mechanism.

Language can directly convey a lack of trust in the veracity of the information. Verbal marking, for example by adding "allegedly", "supposedly", or "some claim", warn that the evidence to back a claim is unreliable, unclear, or missing. We encounter such common expressions of doubt about the veracity of information in gossip-laden conversations, social media communication, in headlines, and on fact-checking sites. In fact, these verbal markers have important journalistic and legal functions to prevent prejudice and wrongful beliefs. Initial evidence from linguistic studies has demonstrated that these expressions can weaken the meaning of information during discourse (Härtl, 2016; Schumacher et al., 2018; Tulling et al., 2021). This weakening can be shown in neurocognitive correlates. Adjectives like "alleged" or "supposed" evoke evaluative processing that is suggested to reflect the need to negate certain aspects of the target information, for example "murderer" (Schumacher et al., 2018). Moreover, a recent study found that factual–but not hypothetical information marked by expressions such as "might"–enhanced neural processing that reflects discourse updating (Tulling et al., 2021). In Study 1 of this dissertation, we show that social-emotional information that is verbally marked respectively, for example "He allegedly bullied his employees", is indeed explicitly judged as less trustworthy compared to information without such markings.

Another important and simple index of news credibility is the media source. Recent large-scale studies have shown that people can well discern the credibility of media sources in service of assessing the veracity of news, and that their layperson credibility ratings well align with those of professional fact-checkers (Allen et al., 2020; Dias et al., 2020; Pennycook & Rand, 2019, 2021). People use resource-efficient and fast heuristics to make such credibility evaluations. Criteria to assess source credibility are cues like familiarity, likability, social endorsement, and reputation, as well as visual layouts (Metzger & Flanagin, 2013). Well-known media sources and their characteristics, including their perceived credibility, are stored in long-term memory. Study 2 and 3 of this dissertation show that participants consistently discern well-known news sources as clearly trustworthy or untrustworthy, for example "Tagesschau" and "Bild" in German media, which have English-speaking analogies such as "BBC" or "The Sun". Moreover, we show that merely the visual layouts of well-known sources' websites can serve as valid cues of trustworthiness.

In sum, verbal expressions as well as source credibility are two frequently used cues of news veracity, theoretically enabling us to mark and identify misinformation as such. It has been suggested that prominent cues of the source are essential to fight the

impact of misinformation and promote truth (e.g., Lorenz-Spreen et al., 2020). However, little is known about how we process the credibility of news when we derive opinions and judgments from it. We tackle this question by investigating the microstructure of information processing with an experimental approach.

1.5 The experimental paradigm of news exposure and derived social judgements

Our general experimental paradigm aims to maintain a balance between an ecologically valid approach and the experimental control necessary for robust neurocognitive evidence (see Fig. 3 for an overview of the general study design; as a variant of a well-established design by e.g., (Abdel Rahman, 2011; Suess et al., 2013).

The paradigm consists of two phases. In Phase 1, participants are exposed to news about a depicted, previously unfamiliar person. The content of news refers to positive, negative, or relatively neutral behaviours of the person, for example "This researcher healed blindness for many". These contents are inspired by authentic news but are revised to gain full control over the presented materials. Thereby this approach also complements large-scale online studies that focus on the identification and spread of misinformation using fully authentic materials (Pennycook & Rand, 2021; Vosoughi et al., 2018). The credibility of news is manipulated either with verbal markers like "allegedly" (Study 1), or in the form of well-known media sources of high or low perceived credibility (Study 2 and 3). As in real life, it is left to the participants to put the information in perspective of its trustworthiness that is easily assessable in form of source credibility or verbal qualification. Crucially, the assignment of news to each face is counterbalanced across participants, meaning that each face is assigned to each combination of positive, negative, and neutral contents in trusted and distrusted news between participants. This allows to precisely measure the neurocognitive correlates and relative contributions of processing the emotional content and the credibility of news while controlling for differences in the appearance of the faces in Phase 2. Additionally, participants rate the likability of the persons before and after news exposure, which served two purposes. First, this rating could be performed spontaneously without any additional information, and it could therefore be compared between pre- and post-exposure and serve as a manipulation check. Second, as person-related information is only indirectly relevant, the results of this task can be compared to the main task of social judgments in Phase 2 where the person-related information is directly task relevant.

In Phase 2, we investigate the effects of the emotional content and the credibility of news by presenting the faces in isolation in a social judgment task. Participants' EEG is acquired, and they explicitly judge the persons based on the news. These social judgments probe a relatively global social evaluation of the person and their social characteristics. Such evaluations are a natural tendency that occurs in real life when we spontaneously form social impressions from visual appearance or semantic information (Bliss-Moreau et al., 2008; Foster, 2004; Todorov et al., 2007; Uhlmann et al., 2015). In real life it is left to us to consider all the available information when we evaluate others. Likewise, we do not explicitly mention the credibility of news when we ask participants to judge the persons. We instructed participants to base their judgments on all the

information from Phase 1, implicitly including the trustworthiness but not explicitly mentioning it.

1.6 Hypotheses for the effects of emotional content and credibility

Based on the theoretical framework and the evidence discussed, specific effects of the emotional content and the credibility of news can be hypothesized for ERPs and social judgments (see Fig. 2 and Fig. 3 Phase 2). Fast brain responses reflected in the EPN are expected to be sensitive to emotion effects mostly regardless of credibility. Thus, enhanced EPN amplitudes for faces associated with negative and positive contents relative to neutral contents should, respectively, be observed for both trustworthy and untrustworthy news. Slow responses reflected in the LPP are expected to take the credibility of the news into account and emotion effects should be modulated accordingly. Therefore, enhanced LPP amplitudes to negative and positive associated contents relative to neutral contents should be found for trustworthy news and, crucially, such modulations should be reduced or absent for untrustworthy news. As the news is put into context of its credibility, the explicit valanced judgments of the persons should be modulated by how trustworthy it is what the news says about them. Consequently, we expect clearly negative and positive judgments for emotional contents compared to neutral contents when news is trustworthy, whereas untrustworthy news should be treated as relatively meaningless and result in much weaker effects of emotional contents.

In addition, the effects of credibility may depend on the valence of the emotional information, such that credibility may primarily modulate the influence of positive, but not negative news content. As discussed earlier, some evidence suggests that negative information might be processed under higher priority and social informational value. As such it is possible that we are prone to neglect the credibility of news specifically in the case of negative information about others.

This dissertation investigates these hypotheses in three studies. Study 1 and 2 aimed to describe the basic neurocognitive underpinnings of our susceptibility to emotional misinformation by investigating verbal marking and source credibility as indicators of news credibility, respectively. Study 3 builds on these results and attempts to intervene with the identified processes, further advancing the understanding of the mechanisms.

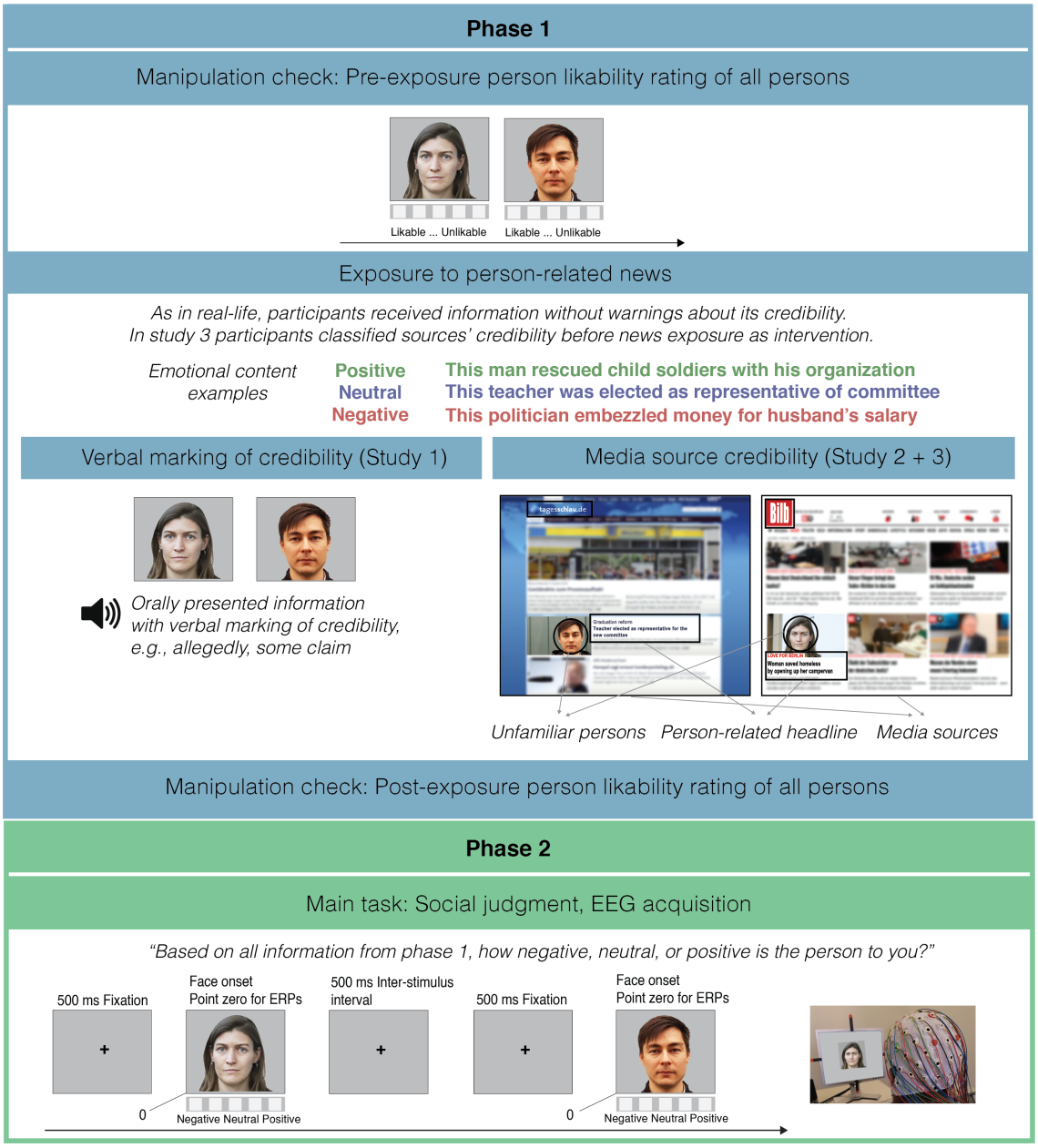

Figure 3. Overview of the general experimental design with two phases and key manipulations. Phase 1 is framed by pre- and post-news-exposure person likability ratings that serve as manipulation check where news is only indirectly relevant. During news exposure, participants are exposed to person-related news about previously unfamiliar people. The emotional content of news was manipulated as either positive, neutral, or negative. Two prominent and commonly used indicators of news credibility are investigated. In Study 1 news is verbally marked as untrustworthy, e.g., by adding "allegedly". In Study 2 and 3 news trustworthiness is manipulated by well-known German media sources that knowingly differ in credibility, e.g., *Tagesschau* and *Bild*. Similar English-speaking sources are for example *BBC News* and *Fox News*. Crucially, the counterbalanced assignment of the face, the emotional content of the person-related information, and the news credibility condition allows to analyse the precise neurocognitive and behavioural correlates of the influence of the manipulations in Phase 2. In Phase 2, that follows Phase 1 after a 15-minute break, the persons are presented in isolation. Participant's EEG is recorded, and they judge each person based on the information from Phase 1.

2. Summary of the present studies

2.1 Study 1: Clear judgments based on unclear evidence: Person evaluation is strongly influenced by untrustworthy gossip

Study 1 (Baum et al., 2018) investigated how we process and consider the verbally communicated lack of trustworthiness (by adding e.g., "allegedly") when we judge persons based on social-emotional information. In Experiment 1 (N=32), neutral or negative news was orally presented while each person was shown on the screen (Phase 1, Fig. 3). Crucially, negative news about the immoral social behaviour of previously unfamiliar persons was either stated as trustworthy fact (e.g., "He bullied his apprentice") or marked as untrustworthy gossip (e.g., "Some claim he bullied his apprentice"). Neutral information was not verbally marked untrustworthy because it sounded ironic or irritating (e.g., "He consulted a technician"). In contrast to the discussed hypotheses, results showed that post-exposure person likability, explicit social judgments, and slow processing reflected in the LPP component are strongly influenced by the negative compared to the neutral news unaffected by the trustworthiness of the information (Phase 2, Fig. 3). A modulation of the EPN was not observed.

Experiment 2 (N=24) aimed to replicate these effects and tested the influence of trustworthiness on positive news. As discussed above, the lack of trustworthiness of non-threatening, positive gossip may be considered while it is neglected for negative gossip to prioritize potentially threatening information. Positive, negative, as well as neutral news was presented, whereby the trustworthiness of emotional information was manipulated like in Experiment 1, e.g., for the untrustworthy positive condition: "He allegedly fought for children's rights over years". Experiment 2 replicated the results of Experiment 1 and extended them to positive gossip. Positive information strongly affected person likability, social judgments, and LPP amplitudes independent of its trustworthiness. EPN modulations suggested a weak modulation for positive contents restricted to 300 to 350 ms, and, if anything, a tendency toward stronger effects for trustworthy information.

The manipulation of credibility in Study 1 reflects the actual use of verbal markers in real-life communication, such as in everyday conversations like gossiping, in news headlines, social media, and on fact checking sites. Moreover, journalistic and legal communication often relies on verbal qualification to prevent consequences of the information, like wrong accusations. Manipulation checks showed that when participants directly rated the trustworthiness of the information, verbally marked gossip was indeed rated as less trustworthy than facts. However, our findings show that understanding that the evidence is unclear does not automatically lead to qualified emotional responses and social judgments. They rather demonstrate a tendency for strong emotional evaluations and judgments even when they are knowingly based on untrustworthy information.

The pattern of effects found in Study 1 contradicts our theoretical predictions. We found highly similar and robust effects when positive or negative social-emotional information was presented fact-like and when it was verbally marked and recognized as

untrustworthy. This contrasts with our expectation that untrustworthy information would be less influential and meaningful, so that slow information processing in the LPP and explicitly derived social judgments would show reduced emotion effects when the information was marked as untrustworthy. At least, we expected to find this theoretically predicted pattern in the case of positive contents when negative contents are prioritized regardless of their trustworthiness as threatening. In conclusion, our findings of emotion effects in both LPP and person judgments regardless of credibility suggest that knowing information is untrustworthy does not automatically result in reconsiderations of emotional appraisals in the service of making accurate and fair social evaluations.

2.2 Study 2: Emotional news affects social judgments independent of perceived media credibility

Study 2 (Baum & Abdel Rahman, 2020) investigated how the credibility we attribute to news sources affects our information processing and the social judgments we derive from news headlines. Building on the results of Study 1, the aim of this preregistered study was to test if theoretically expected effects are found with source credibility as another important, and possibly stronger indication of information trustworthiness. We used real-life well-known news media that are represented in long-term memory along with relatively rich characteristics including their credibility. Indeed, a source credibility rating in the end of the experiment showed that participants hold distinct opinions of how trustworthy they consider the sources. Moreover, this rating confirmed the classification of sources as trusted vs. distrusted that we determined via pre-ratings (N=38) of perceived credibility.

Participants (N=30) were exposed to real news media website layouts including colours and fonts that showed the source logo, and each an experimentally manipulated picture of an unfamiliar person along with a negative, positive, or neutral headline about the person, while all other details were blurred (Phase 1, Fig. 3). We then analyzed how the credibility of the news source is processed and considered when persons are judged based on the news headlines (Phase 2, Fig. 3).

Again in contrast to theory-based expectations, slow processing in the LPP component was dominated by the negative and positive contents of headlines regardless of source credibility. Persons associated with negative or positive headlines were judged more negative or positive than persons associated with neutral headlines regardless of source credibility. Judgments based on emotional contents were further made faster than those based on neutral contents unaffected by source credibility. Person likability showed the same pattern. These findings show once more that emotional news affects late stages of information processing and social judgments regardless of trustworthiness, even when we hold distinct opinions of the sources' credibility.

Fast, emotion- and arousal-related brain responses in the EPN were enhanced for negative contents in news. EPN modulations suggest that this effect of negative contents may be increased for headlines from distrusted sources. While the critical interaction of this modulation was not significant, separate comparisons showed effects only for

distrusted sources. EPN effects by positive contents were not observed. The effects of especially untrustworthy negative news may suggest an enhanced arousal or excitation (cf. Menninghaus et al., 2017). This could result from compounded arousal-related effects by source and headline (cf. Schirmer, 2010).

In sum, we again found a pattern of results in contrast to our theoretical predictions. Like in Study 1, while participants understood the lack of credibility, evaluative brain responses and social judgments were determined by emotional contents independent of their credibility. The similarity of the findings in Study 1 and 2 suggests a general mechanism where emotional contents dominate processing and explicit judgments while source credibility is left unused in the service of reliable social evaluations.

2.3 Study 3: Negative news dominates fast and slow brain responses and social judgments even after source credibility evaluation

Study 3 (Baum & Abdel Rahman, 2021) investigated the malleability of the mechanisms suggested by the findings of Study 1 and 2. The findings of Study 1 and 2 show that social judgments and slow evaluative processing are strongly influenced by emotional contents while being remarkably insensitive to the credibility of the news. As a first attempt to intervene with this bias, the preregistered Study 3 was identical to Study 2 but with one crucial manipulation. As a simple cognitive intervention, participants (N=30) evaluated the credibility of the different news sources *before* they were exposed to news. Participants clearly rated trusted sources as trustworthy, and distrusted sources as untrustworthy. We expected that this active discernment of trustworthiness would enhance the effects of source credibility by making it more salient and increasing its availability when news contents are processed. As such, the intervention tests one aspect of so-called nudging or boosting. Nudging or boosting has been hypothesized to help promote the truth in online discourse (Kozyreva et al., 2020; Lorenz-Spreen et al., 2020), and it has been shown to increase the discernment of news accuracy to some degree (e.g., Pennycook et al., 2020, 2021; Roozenbeek & van der Linden, 2019). However, how nudging may affect the consequences of misinformation for individual information processing and opinions has remained unknown.

We had two alternative expectations regarding if this intervention would affect the results and how. On the one hand, if effective, the intervention should result in theoretically predicted outcomes. This would suggest that increasing the attention to the credibility during news exposure would overcome the dominance of the emotion effects. Specifically, 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 later LPP component, resulting in considerate judgments. The effectiveness of the intervention may differ for positive and negative headline contents, such that source credibility may primarily modulate the influence of positive, but not negative information. On the other hand, if the previously described mechanisms are not readily susceptible to this attempt to intervene, we expected emotional headline contents to strongly affect information processing and social judgments irrespective of source

credibility, in line with the results found in Study 1 and 2. This would suggest that the dominance of emotion is a rather robust effect.

Results showed that this intervention is insufficient to abolish the dominant effects of emotional headlines regardless of source credibility. Participants' social judgments relied to a large extent on the emotional contents of the headlines irrespective of source credibility. In a separate task, we had additionally asked participants to rate how confident they are in their social judgments, and results showed that participants were also highly confident in social judgments based on emotional contents regardless of source credibility.

We also found that this bias might be susceptible to change for positive headline contents, whereas negative social-emotional news seems to be particularly difficult to disregard. Multiple measures suggested signs of source credibility effects for positive contents as hypothesized. These modulations were found in preregistered separate tests for trusted and distrusted sources, finding effects of positive contents only for trusted, but not distrusted sources. These findings are tentative and need future evidence because the specific interactions lacked statistical significance. Specifically, we found that 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. Further in line with predictions, we found LPP effects of positive headline contents only for trusted sources, but not for distrusted sources suggesting that slow brain responses put emotional contents in perspective of their credibility. Furthermore, effects of positive contents in the EPN were also only observed for trusted sources.

In contrast to positive headlines, the effects of negative headlines on fast and slow brain responses and social judgments were immune to the insight of lacking credibility before news exposure. All measures were dominated by negative contents from both trustworthy and untrustworthy sources. Robust effects of negative relative to neutral headline contents regardless of source credibility were observed in fast emotional responses reflected in the EPN, in slow evaluative brain responses reflected in the LPP and in social judgments and their latencies. 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.

The finding that the bias to disregard source credibility may be less pronounced in the positive headline condition was further supported by additional results that explored the role of memory-related processing (please see also section 3.4). We additionally analyzed source credibility effects in the social judgment phase (Phase 2, Fig. 2) by directly considering the individual memories of sources. Specifically, at the end of the experiment participants assigned for each face the source condition that they remembered and associated with the face (we added this task to Study 3 as a first step to address memory-related processing). For the additional analyses we replaced the pre-defined source credibility conditions with those individually remembered. Results showed main effects of individually remembered source credibility in person likability ratings, in social judgments, and in the confidence in judgments. Importantly, the pattern of results confirmed the differential pattern of credibility on positive and negative news headlines.

As in the main results, we found effects of positive headlines for trusted but not for distrusted sources in separate tests (for latencies of judgments and the LPP). These effects were statistically more robust, because we found the critical interactions of positive headline effects with individually assigned source credibility for person likability, confidence in judgments, and the LPP. Remarkably, the influence of associated negative headlines was not modulated by individual source credibility assignments.

Together, these results suggest that it is the dominance of emotional contents, especially negative contents, that weakens the influence of source credibility. This bias is not easily disrupted by evaluating source credibility as a simple preventive intervention. It is particularly difficult to overcome for negative information. Instead, fast and slow brain responses show that emotional headlines bias the information processing system even against better knowledge. Thus, the results of Study 3 further strengthen the evidence alongside Study 1 and 2 for a general mechanism that partially contradicts the predictions derived from the theoretical framework. Besides this key finding, results of Study 3 suggest that the effects of positive contents may be more susceptible to change when trustworthiness is explicitly evaluated before headlines are read. Thus, Study 3 provides initial evidence for leverage against the effects of misinformation.

3. General Discussion

Emotional contents in misinformation can affect mind, brain, and judgments even against better knowledge of its lacking credibility. Neurocognitive evidence reveals that both fast arousal-related processes and slow controlled evaluative processes are influenced by emotional contents regardless of if we trust the news or not. Along the early and later processing stream, it is the emotional contents of news that dominates the information processing in our cognitive system and not the credibility such news holds. The news we hear or read about others affects how positively or negatively we judge them and their social characteristics, even if the news is knowingly untrustworthy.

This is astounding because we expected that our ability to assess, communicate, and understand the credibility of information based on our use of verbal markings as well as on our distinct perceptions of source credibility would guide our news-based opinions and judgments accordingly. Yet, despite this ability, we do not automatically use this knowledge when deriving opinions and judgments about others from person-related news. This pattern of results is highly similar for the two studied and frequently used indicators of how trustworthy news is, verbal marking and source credibility, suggesting a general mechanism of dominant consequences of the emotional contents of news.

This bias even largely prevails a first attempt to intervene, where participants actively discerned the credibility of media sources before being exposed to news and forming opinions. Initial evidence suggests that the bias to neglect credibility can be reduced for positive news contents. Conversely, the results demonstrate the pronounced susceptibility to negative information even from explicitly distrusted sources.

3.1 A general mechanism of the dominant influence of emotional content regardless of credibility

Collectively, the pattern of behavioural and neurocognitive findings suggests a general mechanism of dominant consequences of negative and positive emotional news regardless of its credibility. What does this general mechanism mean considering our expectations based on our theoretical framework and the discussed evidence?

3.1.1 Emotional content drives the fast reception of news

Even the sparse information in emotional headlines–and social misinformation–can impact our information processing rapidly and involuntarily. We found that emotional news, especially with negative contents, influences fast and reflexive brain responses in the EPN. As predicted, this was mostly independent of credibility and mainly reflected the arousal-related influence of emotional contents. Thus, emotional information in news evokes reflexive affective and arousal-related reactions (Barrett & Bar, 2009; Ellsworth & Scherer, 2003), and fast responses to news are driven by this emotional salience.

3.1.2 Emotional content, not credibility, dominates the slow evaluation of relevance and meaning

The emotional content of news receives sustained and elaborate processing, even when of questionable veracity according to its perceived credibility. We found highly robust effects of negative and positive contents in the LPP component, a component that is known to be sensitive to context information and deliberate control (e.g., Blechert et al., 2012; Herbert et al., 2013; Herbert, Pauli, et al., 2011; Rellecke et al., 2012; A Schacht & Sommer, 2009; Schindler et al., 2019, 2020, 2021). This elaborate processing stage was expected to appraise the emotional meaning of news considering its credibility, resulting in reduced emotion effects in response to untrustworthy news (De Houwer et al., 2020b; Ellsworth & Scherer, 2003). Contrary to these expectations, this slow and late stage of information processing weighted the relevance and meaning of news by its emotional contents regardless of its credibility.

In line with the LPP effects, explicit news-based social judgments relied on emotional contents and were remarkably insensitive to the credibility of news. Besides the extremely valenced social judgments, evidence from other tasks further supports this general mechanism. First, likability ratings show the same pattern of dominant emotion effects, suggesting that this mechanism extends to less deliberate social evaluations. Because person likability ratings can be given spontaneously and can be purely based on someone's face without any additional information, news is not directly task relevant. Second, judgments based on emotional contents were made with high confidence irrespective of news credibility. This shows a reliance on emotional content even when asked for the subjective confidence in social judgments, an evaluation that is assumed to be more deliberate and metacognitive than the main social judgment task (Petty et al., 2007).

This general mechanism of dominant emotion effects is even relatively robust against an intervention aimed to increase the attention to credibility. Specifically, participants explicitly evaluated the credibility of media sources before they were exposed to their headlines (Study 3). This should have strengthened the saliency and availability of credibility and by that the influence on slow evaluation and judgments. Although participants clearly discerned the trustworthiness of media sources before they read the news, this simple intervention was insufficient to abolish the bias, at least for negative news (see discussion below). This implies that the dominance of emotion is not simply explained by gross inattention to credibility (see also section 3.2) but further suggests that emotion outweighs credibility.

The described mechanism contrasts the assumption that common and meaningful indicators of credibility in everyday life protect against the impact of misinformation. Instead, they seem strikingly inconsequential when news has emotional content. We tested two prominent indicators of how trustworthy or untrustworthy news is, that is distinct perceptions of media source credibility and frequently used verbal expressions of lacking evidence, e.g., "allegedly". The similarity of the findings despite two different indicators of news credibility further speaks for a general mechanism. The knowledge about communicating and assessing credibility is not automatically leveraged against the influence of misinformation and does not directly work in service of fair and accurate social judgments.

Taken together, our findings suggest that emotionally evocative content dominates the intrinsic meaning and relevance of the news, resulting in social judgments reliant on these contents. In contrast to the theoretical expectation, credibility is not the key determinant of how meaningful and relevant social-emotional news is evaluated and how persons are judged based on news, even when credibility is explicitly evaluated before news exposure.

3.1.2 Credibility effects might depend on the valence of contents

Additional to the many similarities of the effects of negative and positive news regardless of credibility, we also find indications for differential effects of credibility. Specifically, thinking about source credibility before news exposure might protect us against the influence of positive misinformation to some degree. After the cognitive intervention in Study 3, we found a pattern of results that suggests that the bias of dominant emotion effects is easier to overcome for positive news, while negative social-emotional information seems to be particularly difficult to disregard, even when it stems from sources one explicitly evaluates as untrustworthy. We had hypothesized that the theoretically expected effects of credibility may be found at least for positive news if the credibility if more readily neglected in case of threatening negative information. Indeed, the results showed effects of positive news in slow processing and judgment latencies only for trusted sources, but not for distrusted sources. Nevertheless, the intervention was not sufficient to lead to significantly more neutral explicit judgments of persons associated with distrusted positive news, calling for future research of more effective strategies against the effects of misinformation (please see section 3.4).

Negative misinformation seems hard to disregard despite it knowingly lacking credibility. This dominance might weaken the influence of credibility. Negative information may be processed under high priority as a protection against the potential threat an allegedly dangerous person could pose (Öhman & Mineka, 2001). This preferential processing seems to come at the cost of critical thinking. Recent evidence supports this potential explanation by showing that negative emotional narrative contexts reduce critical semantic analyses (Aristei et al., 2022). In comparison, one could speculate that positive emotional information may be related to a broadened cognitive scope (Fredrickson, 2001; Zimmerman & Kelley, 2010). For example, positive emotions can sometimes enhance the processing of contexts (Madan et al., 2019; Talarico et al., 2009).

The current findings cannot determine if the differential patterns for positive and negative headlines are related to differences in valence or in intensity or both. However, pre-ratings of our materials show that the arousal does not differ between positive and negative contents. Additional control analyses further suggest that valence may play an independent role because differences in ERPs in response to positive vs. negative headlines were only found for distrusted sources, and not for trusted sources. The exact mechanisms underlying differential effects of credibility on positive and negative news after intervention are a topic for future research (see section 3.4).

3.1.4 What social misinformation means for person perception and judgments

The news we hear and read about others influences how we perceive their face and evaluate their social characteristics, and as shown here, even when this person knowledge is merely based on untrustworthy, possibly false information.

We show that social misinformation can influence how we perceive persons as reflected in earlier and later brain responses. These findings suggest that emotional information comes online relatively early and reflexively in the face processing stream and dominates also later, post-perceptual processing (c.f., Abdel Rahman, 2011; Schindler et al., 2021; Suess et al., 2014; Wieser et al., 2014; Wieser & Brosch, 2012). N170 effects as observed in Study 3 show that even earlier visual processing of the face that is related to structural encoding is affected by negative news (Giménez-Fernández et al., 2020, 2021; Krasowski et al., 2021; Luo et al., 2016; Schindler et al., 2021; Xu et al., 2016). These findings are in line with models of face perception that assume that person knowledge can already influence early and reflexive processing of faces (Haxby et al., 2000; Gobbini & Haxby, 2007).

This influence is found even though we could rely more on the information that is directly observable from the faces, for example their (neutral) facial expressions, when the person-related knowledge we have available cannot be trusted. In general, prior knowledge can be highly useful to navigate the social world via rapid top-down effects or predictions on our perception as well as early and later information processing (Abdel Rahman, 2011; FeldmanHall & Shenhav, 2019; Maier et al., 2022; Wieser & Brosch, 2012). This helps to quickly reduce uncertainty about other's intentions or to understand facial expressions. Indeed, the effects of social misinformation are comparable to effects found for affective biographical knowledge about well-known faces (Abdel Rahman, 2011, Suess et al., 2014). To gain biographical knowledge one has likely encountered

many news reports and has built up a possibly more valid knowledge. Here, top-down effects were found also by potential misinformation that should have little epistemic meaning.

To make accurate and fair person evaluations, later evaluative processes and explicit social judgments could put initial arousal-related responses into perspective of the credibility of the information. However, we find that we rely more on the emotional contents than the credibility of the information, suggesting that we care little about the credibility, either deliberately or unintentionally. This implies that we might constrain our social evaluations to the emotional substance and its potential relevance for our own well-being and coping possibilities even at the risk of forming misleading or inaccurate judgments of others, for example lost opportunities for social relations and cooperation, and potential damage of existing relationships.

3.2 Experimental neurocognitive evidence underlying real-life behaviour

We translated many aspects of real-life media encounters into our design, but one may suggest that it nevertheless differs from natural situations in many ways. That is true, yet we took additional measures to show that our experimental approach unlikely caused the pattern of results. In fact, we are confident that our approach represents a unique strength because it enables us to obtain robust evidence of the microstructure of information processing. This is a crucial foundation to be able to vary from controlled designs in the future and it complements other research in the field that has largely focused on large-scale observational and behavioural studies using completely original but uncontrolled materials.

Was our manipulation of news credibility reasonable and did participants understand it? We used indicators of news credibility that are common in real life, and we can show that these are understood. Verbal markers such as *allegedly* are frequently used in gossip-laden conversations, social media communication, or news headlines, moreover such expressions have important legal and journalistic functions of preventing prejudice and wrongful accusations. Well-known media sources and original layouts closely resemble news media encountered in real life that are stored in long-term memory including their credibility. We repeatedly show that people understand these indicators and can discern trustworthiness from these cues. Participants in pre-ratings as well as in experimental samples a) differentiate the trustworthiness of verbally marked information from un-marked information and b) differentiate sources in trustworthiness ratings (replicating other findings, e.g., Metzger & Flanagin, 2013; Pennycook & Rand, 2019). Furthermore, short interviews after the experiments showed that participants did not generally doubt the authenticity of the media reports, supporting our approach to use real-life media stimuli.

Did participants notice the difference in credibility? We are confident that the credibility of media sources was noticed during news exposure for several reasons. First, each news website presentation started with showing the prominent source logo alone for one second before the face and headline information was presented, giving extra time to process source information. Second, we demonstrated in a control study via tracking

active eye movements that participants sample source information during the news exposure. Third, in the same control study we showed that blurred website layouts with retracted source names are reliable cues for credibility. Finally, the intervention of explicitly classifying sources before news exposure should have even increased the attention to the sources.

Thus, we are confident that the findings reflect processes that happen in real life. It is possible that we notice indicators of credibility and grasp their meaning, but just like in real life, we may not fully engage with or encode this information. Such potential mechanisms should be targeted by interventions against the influence of misinformation in future research (see section 3.4).

The experimental situation is somewhat artificial due to the experimentally necessary quantity of faces and repetitions in the judgment task. However, we can show that it is unlikely that this has caused the results. To address this concern, we show that post hoc analyses accounting for task repetitions and tests of only first judgments reveal the same pattern of results. Furthermore, we show in a control experiment that the same pattern of dominant effects of emotional information regardless of credibility is found when participants make immediate, one-time judgments after each news is presented. Thus, removing repetitions in both exposure and judgments does not change the pattern of results.

Finally, the tendency for effects of source credibility after the intervention for positive news, but not negative news, further supports the conclusion that our results show a general mechanism that is not simply due to the experimental procedure. Overall, we are confident that the repeatedly found pattern of effects, our control experiments and analyses, plus the differential pattern found after a first intervention imply a non-trivial general mechanism as discussed and described in the previous section (3.1).

3.3 On advancing the understanding of the susceptibility to misinformation

Our insights on the consequences of misinformation on the level of the individual's cognitive system complement research that has focused on people's knowledge and ability to discern news credibility. This research field is very active and has developed in parallel to the timeline of this dissertation. The results of our manipulation checks are in line with research demonstrating that people can distinguish the credibility of media sources and that they understand the meaning of verbally communicated lack of trustworthiness (e.g., Metzger & Flanagin, 2013; Pennycook & Rand, 2019; Tulling et al., 2021; van der Bles et al., 2020). However, we show that social-emotional misinformation affects our judgments even against this better knowledge. Our neurocognitive evidence precisely shows what cognitive processes are affected and when. Thus, our findings suggest that the ability to identify credibility is not automatically protective against the impact of misinformation. In real-life we are not usually directly asked to identify misinformation or prompted to think about the credibility of news. Often, we are just confronted with news, process it, form impressions, and derive judgments. Moreover, we show that the general tendency to rely more on news' evoked

emotions than its credibility even holds when known sources are explicitly evaluated before news is read, at least when news has negative emotional contents.

This is further relevant to research on interventions against misinformation that focus on helping people to identify information as untrustworthy, specifically nudging, boosting, and inoculation interventions. Such interventions aim to prompt or educate people with the aim to reduce their susceptibility to misinformation (e.g., Fazio, 2020; Kozyreva et al., 2020; Lorenz-Spreen et al., 2020; Pennycook et al., 2020, 2021; Roozenbeek & van der Linden, 2019). For example, it has been proposed to highlight the news source and to nudge people to pay attention to cues that indicate the quality of news to curb the influence of misinformation and promote trustworthy news (Lorenz-Spreen et al., 2020). Recent empirical tests of nudging interventions asked participants to think about the accuracy of one headline on a general topic (Pennycook et al., 2021, 2020). Subsequently, verbatim headlines from various sources were presented and the intervention improved participants' explicit differentiation of accurate and inaccurate headlines related to political and health information. Our findings complement this research on different levels. First, our results suggest that the merit of interventions needs to be further investigated in the consequences of news exposure, even if interventions make people aware of differences in credibility. Second, the neurocognitive evidence of such consequences adds a perspective that has been lacking but can identify how key cognitive processes respond to interventions. Third, the general mechanisms we described show that the effectiveness of interventions for influencing different outcomes may depend on the context, specifically the emotional content of news and its valence. These insights directly motivate future research on possible interventions (see section 3.4).

Diverse disciplines that study misinformation from different angles can derive testable hypotheses from our experimental evidence that has identified emotional contents as a key determinant of how we process misinformation. For instance, additional sentiment analyses of the verbatim headlines used in recent studies (e.g., Pennycook & Rand, 2018) shows that negative rather than positive contents were more often accepted as correct (Borukhson et al., 2021). However, the emotional contents of the headlines were not considered in the original study.

The described mechanisms help explain phenomena in real-life news communication. In massive online communication, we are confronted with contents that evoke emotions such as moral outrage at a high rate. Emotional contents can powerfully capture our attention, are more readily shared with others, and are therefore often favoured in online media where attention is an economically valuable resource (Bavel et al., 2020; Brady et al., 2017, 2020; Crockett, 2017; Schöne et al., 2021; Vosoughi et al., 2018). For example, the social media platform Facebook profits most of content that evokes anger and negative emotions, and even charges less for ads with negative than with positive content (*Facebook Whistleblowerin Frances Haugen Im Talk Über Die Facebook Papers | ZDF Magazin Royale - YouTube*, 2021). Our findings suggests that this makes sense in terms of maximizing impact: Emotional–especially negative–content has an even bigger impact than expected, because it is relevant to our thinking and judgments even when it is lacking credibility.

The minor role of the credibility of emotional messages, as shown here, gives misinformation an advantage and helps explain its "success" on large scales. For example, negative false news evoking emotions like fear and disgust spreads worryingly quickly and widely in online media (Vosoughi et al., 2018). These large-scale propagation measures of news may even boil down to mechanisms of how individuals respond to news. Latest research of the spreading patterns indeed suggests that the "success" of misinformation is best explained by its person-to-person infections (Juul & Ugander, 2021). Our findings of behavioural and neurocognitive responses to emotional misinformation on the individual level help reveal what these mechanisms are.

3.4 Limitations and future research

Going forward, one important goal for future research is to understand *why* the influence of emotional content is so dominant and the role of credibility so minor. To help explain this bias, several targets of future research directly arise from the current work. First, one factor in the persistence of this bias is related to the valence and intensity of the emotional contents, with negative contents prevailing our attempt to intervene more strongly than positive contents. The effects of the emotional intensity of news, and the relative contribution of valence should be targeted in future research. Second, mechanisms of learning and memory may be related to the pattern of results. We show that participants notice and process the different sources and their credibility during news exposure (section 3.2). Additional findings suggest that memory may play a role for main effects of news credibility, but generally corroborate that the bias of emotional dominance may be overcome for positive news but prevails for negative news. Future research is needed to investigate the relative contribution of learning and memory related mechanisms directly. Third, modulations of fast brain responses by credibility could tell us more about certain mechanisms. Although theoretically not expected, results of Study 2 indicated that the EPN effect of negative headlines was most pronounced for distrusted sources. However, this modulation was not replicated in Study 3 which employed an intervention and showed clear and robust EPN effects for negative news regardless of credibility. It can be speculated that untrustworthy negative social-emotional news may induce pleasurable states of enhanced arousal or excitation (cf. Menninghaus et al., 2017), increasing the impact of negative information (Kahneman & Tversky, 1979; Zillmann, 2013). Future research should investigate the hypothesis that the popularity and sensational pleasure of gossip and misinformation may be related to those effects. Furthermore, after the intervention of Study 3, EPN effects for positive news were found only when news came from trustworthy sources, but not when it came from distrusted sources. Such early modulations may suggest that interventions can influence early effects of emotional contents on information processing. Future research should test this hypothesis, since finding out what aspects of interventions affect also fast brain responses can shed light on the mechanisms of an intervention. Fourth, meta-cognitive processes like the confidence in social judgments or the fluency with which judgments are derived may help explain the bias (Brashier & Marsh, 2020; Wang et al., 2016). In Study 3 we found that social judgments based on emotional contents are made with high confidence

regardless of credibility. Future research should directly investigate mental processes related to meta-cognition, for example cognitive resource allocation. Fifth, previous knowledge and other priors might play important roles in news-based social judgments. In the current work, news featured previously unfamiliar persons as it can often be the case in news coverage. However, it is of course also common that news is about famous persons or familiar acquaintances. It is possible that previous knowledge about a specific person can modulate the influence and relevance of untrustworthy information, especially when the news is contradicting previous impressions (e.g., Cone et al., 2019; Mende-Siedlecki et al., 2013; Rapp, 2016). Overall, future research into the mechanisms of the bias can employ other tasks and (neurocognitive) measures that target different processes. For example, the effects of news could be tested more implicitly in comparison to the social judgments employed here, or they could be tested for decisions that require more deliberation such as punishments or rewards. Finally, individual differences in the susceptibility to emotional misinformation may help to further explain why emotional contents have a dominant effect. For example, people over the age of 65 have shared more misinformation than younger people on social media and through messaging apps (e.g., Brashier & Schacter, 2020), making age-related cognitive processes another interesting target for future investigations.

Another important goal of future research is to use the current insights to identify strategies to reduce the influence of misinformation and consider the veracity of news. Previous strategies have often not had the wanted success, and one reason is that they have not taken into account basic cognitive, social, and emotional mechanisms, into which insight has been lacking (Ecker et al., 2022). These strategies have for example attributed people's susceptibility to misinformation to a lack of information that can be overcome by providing more facts (Lewandowsky et al., 2017; Simis et al., 2016). Based on the neurocognitive mechanisms shown here, we can investigate the aspects of the cognitive system that must be targeted to successfully intervene with the described bias. The current findings show that a simple intervention focused on contemplating source credibility might be able to reduce the bias for positive headlines but seems insufficient to reduce it for negative headlines. One possibility to move forward is to target the dominance of emotional contents more directly. Employing emotion regulation seems to be a promising avenue. For example, we could deliberately choose not to know about certain news or shield ourselves from exposure (Hertwig & Engel, 2016). However, it is not always possible and sometimes even impossible to avoid being exposed to misinformation in today's massive online communication landscape. Online media prominently displays emotional contents that can powerfully capture our attention and it is unlikely that online media will be free from misinformation (e.g., Kozyreva et al., 2020). Therefore, it is important to develop other strategies that build on our understanding of neuro-cognitive mechanisms and specifically protect the individual information processing system and judgments. Such strategies include, for example, educating about the consequences of social-emotional misinformation for us individually and as social beings, and by employing cognitive reappraisal of untrustworthy emotional contents to reduce their impact (e.g., Gross, 2015; Maroney & Gross, 2014; Wang et al., 2021; Webb et al., 2012).

3.5 Final remarks

In conclusion, understanding the lack of credibility matters little for the consequences of emotional contents in social misinformation affecting mind, brain, and judgments. It is the emotional content that matters. Neurocognitive evidence reveals that emotional content affects fast-reflexive emotional responses and dominates slow-controlled processing regardless of credibility. This bias is especially robust for negative contents, even when attention to credibility is increased. These insights shed light on the cognitive mechanisms underlying the processing and impact of social misinformation on the individual level. Thus, they help explain the "success" of misinformation on large scales and identify key targets for future interventions. Ultimately, a precise understanding of our cognitive-emotional responses, vulnerabilities, and strengths can empower us to protect ourselves against misinformation biasing our minds and behaviours.

References

Abdel Rahman, R. (2011). Facing Good and Evil: Early Brain Signatures of Affective Biographical Knowledge in Face Recognition. *Emotion*, *11*(6), 1397–1405. https://doi.org/10.1037/a0024717

Allen, J., Howland, B., Mobius, M., Rothschild, D., & Watts, D. J. (2020). Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances*, *6*(14), eaay3539. https://doi.org/10.1126/sciadv.aay3539

Anderson, E., Siegel, E. H., Bliss-Moreau, E., & Barrett, L. F. (2011). The visual impact of gossip. *Science*, *332*(6036), 1446–1448. https://doi.org/10.1126/science.1201574

Aristei, S., Knoop, C. A., Lubrich, O., Nehrlich, T., Enge, A., Stark, K., Sommer, W., & Abdel Rahman, R. (2022). Affect as Anaesthetic. How emotional contexts modulate the processing of counterintuitive concepts. *In Press*.

Aviezer, H., Ensenberg, N., & Hassin, R. R. (2017). The inherently contextualized nature of facial emotion perception. *Current Opinion in Psychology*, *17*, 47–54. https://doi.org/10.1016/j.copsyc.2017.06.006

Barrett, L. F., & Bar, M. (2009). See it with feeling: affective predictions during object perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1325–1334. https://doi.org/10.1098/rstb.2008.0312

Baum, J., & Abdel Rahman, R. (2020). Emotional news affects social judgments independent of perceived media credibility. *Social Cognitive and Affective Neuroscience*, *16*(3), 280–291. https://doi.org/10.1093/scan/nsaa164

Baum, J., & Abdel Rahman, R. (2021). Negative news dominates fast and slow brain responses and social judgments even after source credibility evaluation. *NeuroImage*, *244*(August), 118572. https://doi.org/10.1016/j.neuroimage.2021.118572

Baum, J., Rabovsky, M., Rose, S. B., & Abdel Rahman, R. (2018). Clear Judgments Based on Unclear Evidence: Person Evaluation Is Strongly Influenced by Untrustworthy Gossip. *Emotion*, *20*(2), 248–260. https://doi.org/10.1037/emo0000545

Bavel, J. J. V., 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 Behaviour*, *4*(5), 460–471. https://doi.org/10.1038/s41562-020-0884-z

Bhadani, S., Yamaya, S., Flammini, A., Menczer, F., Ciampaglia, G. L., & Nyhan, B. (2022). Political audience diversity and news reliability in algorithmic ranking. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-021-01276-5

Blechert, J., Sheppes, G., Tella, C., Williams, H., & Gross, J. J. (2012). See What You Think:

Reappraisal Modulates Behavioral and Neural Responses to Social Stimuli. *Psychological Science*, *23*(4), 346–353. https://doi.org/10.1177/0956797612438559

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. https://doi.org/10.1037/1528-3542.8.4.479

Borukhson, D., Lorenz-Spreen, P., & Ragni, M. (2021). When Does an Individual Accept Misinformation? *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*(43). https://doi.org/https://escholarship.org/uc/item/75b1q7r4

Brady, W. J., Gantman, A. P., & Van Bavel, J. J. (2020). Attentional capture helps explain why moral and emotional content go viral. *Journal of Experimental Psychology: General*, *149*(4), 746–756. https://doi.org/10.1037/xge0000673

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., Van Bavel, J. J., & Fiske, S. T. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(28), 7313–7318. https://doi.org/10.1073/pnas.1618923114

Brashier, N. M., & Marsh, E. J. (2020). Judging Truth. *Annual Review of Psychology*, *71*(1), 499–515. https://doi.org/10.1146/annurev-psych-010419-050807

Brashier, N. M., & Schacter, D. L. (2020). Aging in an Era of Fake News. *Current Directions in Psychological Science*, *29*(3), 316–323. https://doi.org/10.1177/0963721420915872

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*(3), 305–327. https://doi.org/10.1111/j.2044-8295.1986.tb02199.x

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1999). The affect system has parallel and integrative processing components: Form follows function. *Journal of Personality and Social Psychology*, *76*(5), 839–855. https://doi.org/10.1037/0022-3514.76.5.839

Cone, J., Flaharty, K., & Ferguson, M. J. (2019). Believability of evidence matters for correcting social impressions. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(20), 9802–9807. https://doi.org/10.1073/pnas.1903222116

Crockett, M. J. (2017). Moral outrage in the digital age. *Nature Human Behaviour*, *1*(11), 769–771. https://doi.org/10.1038/s41562-017-0213-3

Cunningham, W. A., & Zelazo, P. D. (2007). Attitudes and evaluations: a social cognitive neuroscience perspective. *Trends in Cognitive Sciences*, *11*(3), 97–104. https://doi.org/10.1016/j.tics.2006.12.005

De Houwer, J., Van Dessel, P., & Moran, T. (2020a). Attitudes beyond associations: On the role of propositional representations in stimulus evaluation. In *Advances in Experimental Social Psychology* (Vol. 61, pp. 127–183). Elsevier. https://doi.org/10.1016/bs.aesp.2019.09.004

De Houwer, J., Van Dessel, P., & Moran, T. (2020b). Attitudes beyond associations: On the role

of propositional representations in stimulus evaluation. In B. Gawronski (Ed.), *Advances in Experimental Social Psychology* (Vol. 61, pp. 127–183). Academic Press. https://doi.org/10.1016/bs.aesp.2019.09.004

Dias, N., Pennycook, G., & Rand, D. G. (2020). Emphasizing publishers does not effectively reduce susceptibility to misinformation on social media. *Harvard Kennedy School Misinformation Review*, 1–12. https://doi.org/10.37016/mr-2020-001

Dunbar, R. I. M. (2004). Gossip in evolutionary perspective. *Review of General Psychology*, *8*(2), 100–110. https://doi.org/10.1037/1089-2680.8.2.100

Dunbar, R. I. M., Marriott, A., & Duncan, N. D. C. (1997). Human conversational behavior. *Human Nature*, *8*(3), 231–246. https://doi.org/10.1007/BF02912493

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

Ehret, P. J., Monroe, B. M., & Read, S. J. (2015). Modeling the Dynamics of Evaluation: A Multilevel Neural Network Implementation of the Iterative Reprocessing Model. *Personality and Social Psychology Review*, *19*(2), 148–176. https://doi.org/10.1177/1088868314544221

Eiserbeck, A., & Abdel Rahman, R. (2020). Visual consciousness of faces in the attentional blink: Knowledge-based effects of trustworthiness dominate over appearance-based impressions. *Consciousness and Cognition*, *83*, 102977. https://doi.org/10.1016/j.concog.2020.102977

Eiserbeck, A., Enge, A., Rabovsky, M., & Abdel Rahman, R. (2021). Distrust before first sight: Knowledge- and appearance-based effects of trustworthiness on the visual consciousness of faces. *BioRxiv*. https://doi.org/https://doi.org/10.1101/2021.02.24.432562

Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Series in Affective Science.* (pp. 572–595). New York, NY, US: Oxford University Press.

Facebook Whistleblowerin Frances Haugen im Talk über die Facebook Papers | ZDF Magazin Royale - YouTube. (2021). https://www.youtube.com/watch?v=ws06adOKNUk&ab_channel=ZDFMAGAZINROYALE

Fazio, L. K. (2020). Recognizing the Role of Psychological Science in Improving Online Spaces. *Psychological Science in the Public Interest*, *21*(3), 99–102. https://doi.org/10.1177/1529100620972100

FeldmanHall, O., & Shenhav, A. (2019). Resolving uncertainty in a social world. *Nature Human Behaviour*, *3*(5), 426–435. https://doi.org/10.1038/s41562-019-0590-x

Ferrari, C., Oh, D. W., Labbree, B. P., & Todorov, A. (2020). Learning the affective value of people: More than affect-based mechanisms. *Acta Psychologica*, *203*, 103011. https://doi.org/10.1016/j.actpsy.2020.103011

Fields, E. C., & Kuperberg, G. R. (2012). It's All About You: An ERP Study of Emotion and Self-Relevance in Discourse. *NeuroImage*, *62*(1), 562–574. https://doi.org/10.1016/j.neuroimage.2012.05.003

Foster, E. K. (2004). Research on gossip: Taxonomy, methods, and future directions. *Review of General Psychology*, *8*(2), 78–99. https://doi.org/10.1037/1089-2680.8.2.78

Fredrickson, B. L. (2001). The Role of Positive Emotions in Positive Psychology. *The American Psychologist*, *56*(3), 218–226. https://doi.org/10.1037/0003-066x.56.3.218

Galli, G., Feurra, M., & Viggiano, M. P. (2006). "Did you see him in the newspaper?" Electrophysiological correlates of context and valence in face processing. *Brain Research*, *1119*(1), 190–202. https://doi.org/10.1016/j.brainres.2006.08.076

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692–731. https://doi.org/10.1037/0033-2909.132.5.692

Giménez-Fernández, T., Fernández-Folgueiras, U., Fondevila, S., Méndez-Bértolo, C., García-Rubio, M. J., Hernández-Lorca, M., Kessel, D., & Carretié, L. (2021). Enhanced N170 to outgroup faces: Perceptual novelty or prejudice? *Social Neuroscience*, *16*(3), 252–264. https://doi.org/10.1080/17470919.2021.1889658

Giménez-Fernández, T., Kessel, D., Fernández-Folgueiras, U., Fondevila, S., Méndez-Bértolo, C., Aceves, N., García-Rubio, M. J., & Carretié, L. (2020). Prejudice drives exogenous attention to outgroups. *Social Cognitive and Affective Neuroscience*, *15*(6), 615–624. https://doi.org/10.1093/scan/nsaa087

Gobbini, M. I., & Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia*, *45*(1), 32–41. https://doi.org/10.1016/j.neuropsychologia.2006.04.015

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, *106*(1), 148–168. https://doi.org/10.1037/a0034726

Gross, J. J. (2015). Emotion Regulation: Current Status and Future Prospects. *Psychological Inquiry*, *26*(1), 1–26.

Hajcak, G., Macnamara, A., & Olvet, D. M. (2010). Event-related potentials, emotion, and emotion regulation: An integrative review. *Developmental Neuropsychology*, *35*(2), 129–155. https://doi.org/10.1080/87565640903526504

Härtl, H. (2016). *Sogenannt ('so-called') and the lexicon-pragmatics interface*. https://doi.org/https://ling.auf.net/lingbuzz/003017/current.pdf

Haxby, J. V, Hoffman, E. A., & Gobbini, M. I. (2000). 08_Haxbytics2000. *Trends Cogn Sci.*

4:223–233, *4*(6), 223–233.

Herbert, C., Herbert, B. M., Ethofer, T., & Pauli, P. (2011). His or mine? The time course of self-other discrimination in emotion processing. *Social Neuroscience*, *6*(3), 277–288. https://doi.org/10.1080/17470919.2010.523543

Herbert, C., Pauli, P., & Herbert, B. M. (2011). Self-reference modulates the processing of emotional stimuli in the absence of explicit self-referential appraisal instructions. *Social Cognitive and Affective Neuroscience*, *6*(5), 653–661. https://doi.org/10.1093/scan/nsq082

Herbert, C., Sfaerlea, A., & Blumenthal, T. (2013). Your emotion or mine: Labeling feelings alters emotional face perception- An ERP study on automatic and intentional affect labeling. *Frontiers in Human Neuroscience*, *7*(JUL). https://doi.org/10.3389/fnhum.2013.00378

Hertwig, R., & Engel, C. (2016). Homo Ignorans: Deliberately Choosing Not to Know. *Perspectives on Psychological Science*, *11*(3), 359–372. https://doi.org/10.1177/1745691616635594

Hinojosa, J. A., Mercado, F., & Carretié, L. (2015). N170 sensitivity to facial expression: A meta-analysis. *Neuroscience and Biobehavioral Reviews*, *55*, 498–509. https://doi.org/10.1016/j.neubiorev.2015.06.002

Junghöfer, M., Bradley, M. M., Elbert, T. R., & Lang, P. J. (2001). Fleeting images: A new look at early emotion discrimination. *Psychophysiology*, *38*(2), 175–178. https://doi.org/10.1017/S0048577201000762

Junghöfer, M., Rehbein, M. A., Maitzen, J., Schindler, S., & Kissler, J. (2017). An evil face? Verbal evaluative multi-CS conditioning enhances face-evoked mid-latency magnetoencephalographic responses. *Social Cognitive and Affective Neuroscience*, *12*(4), 695–705. https://doi.org/10.1093/scan/nsw179

Juul, J. L., & Ugander, J. (2021). Comparing information diffusion mechanisms by matching on cascade size. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(46). https://doi.org/10.1073/pnas.2100786118

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American Psychologist*, *58*(9), 697–720.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, *47*(2), 263.

Kissler, J., & Strehlow, J. (2017). Something always sticks? how emotional language modulates neural processes involved in face encoding and recognition memory. *Poznan Studies in Contemporary Linguistics*, *53*(1), 63–93. https://doi.org/10.1515/psicl-2017-0004

Klein, F., Iffland, B., Schindler, S., Wabnitz, P., & Neuner, F. (2015). This person is saying bad things about you: The influence of physically and socially threatening context information on the processing of inherently neutral faces. *Cognitive, Affective and Behavioral*

Neuroscience, *15*(4), 736–748. https://doi.org/10.3758/s13415-015-0361-8

Kozyreva, A., Lewandowsky, S., & Hertwig, R. (2020). Citizens Versus the Internet: Confronting Digital Challenges With Cognitive Tools. *Psychological Science in the Public Interest*, *21*(3), 103–156. https://doi.org/10.1177/1529100620946707

Krasowski, C., Schindler, S., Bruchmann, M., Moeck, R., & Straube, T. (2021). Electrophysiological responses to negative evaluative person-knowledge: Effects of individual differences. *Cognitive, Affective and Behavioral Neuroscience*, *21*(4), 822–836. https://doi.org/10.3758/s13415-021-00894-w

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news: Addressing fake news requires a multidisciplinary effort. *Science*, *359*(6380), 1094–1096. https://doi.org/10.1126/science.aao2998

Lewandowsky, S., Ecker, U. K. H., Cook, J., & States, U. (2017). Understanding and Coping with the "Post-Truth" Era. *Journal of Applied Research in Memory and Cognition*, *6*(4), 353–369. http://dx.doi.org/10.1016/j.jarmac.2017.07.008

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest, Supplement*, *13*(3), 106–131. https://doi.org/10.1177/1529100612451018

Lieberman, M. D. (2007). Social cognitive neuroscience: A review of core processes. *Annual Review of Psychology*, *58*(1), 259–289. https://doi.org/10.1146/annurev.psych.58.110405.085654

Lieberman, M. D., Gaunt, R., Gilbert, D. T., & Trope, Y. (2002). Reflexion and reflection: A social cognitive neuroscience approach to attributional inference. *Advances in Experimental Social Psychology*, *34*, 199–249. https://doi.org/10.1016/s0065-2601(02)80006-5

Lorenz-Spreen, P., Lewandowsky, S., Sunstein, C. R., & Hertwig, R. (2020). How behavioural sciences can promote truth, autonomy and democratic discourse online. *Nature Human Behaviour*, *4*(11), 1102–1109. https://doi.org/10.1038/s41562-020-0889-7

Luo, Q. L., Wang, H. L., Dzhelyova, M., Huang, P., & Mo, L. (2016). Effect of affective personality information on face processing: Evidence from ERPs. *Frontiers in Psychology*, *7*(MAY), 1397. https://doi.org/10.3389/fpsyg.2016.00810

Madan, C. R., Scott, S. M. E., & Kensinger, E. A. (2019). Positive emotion enhances association-memory. *Emotion*, *19*(4), 733–740. https://doi.org/10.1037/emo0000465

Maier, M., Blume, F., Bideau, P., Hellwich, O., & Abdel Rahman, R. (2022). Knowledge-augmented face perception: Prospects for the Bayesian brain-framework to align AI and

human vision. *Consciousness and Cognition*, *101*, 103301. https://doi.org/10.1016/J.CONCOG.2022.103301

Maroney, T. A., & Gross, J. J. (2014). The ideal of the dispassionate judge: An emotion regulation perspective. *Emotion Review*, *6*(2), 142–151. https://doi.org/10.1177/1754073913491989

Mattarozzi, K., Todorov, A., & Codispoti, M. (2015). Memory for faces: the effect of facial appearance and the context in which the face is encountered. *Psychological Research*, *79*(2), 308–317. https://doi.org/10.1007/s00426-014-0554-8

Mende-Siedlecki, P., Cai, Y., & Todorov, A. (2013). The neural dynamics of updating person impressions. *Social Cognitive and Affective Neuroscience*, *8*(6), 623–631. https://doi.org/10.1093/scan/nss040

Menninghaus, W., Wagner, V., Hanich, J., Wassiliwizky, E., Jacobsen, T., & Koelsch, S. (2017). The Distancing-Embracing model of the enjoyment of negative emotions in art reception. *Behavioral and Brain Sciences*, *40*, 26. https://doi.org/10.1017/S0140525X17000309

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

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review*, *108*(3), 483–522. https://doi.org/10.1037/0033-295X.108.3.483

Otten, M., Seth, A. K., & Pinto, Y. (2017). A social Bayesian brain: How social knowledge can shape visual perception. *Brain and Cognition*, *112*, 69–77. https://doi.org/10.1016/j.bandc.2016.05.002

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, *592*(7855), 590–595. https://doi.org/10.1038/s41586-021-03344-2

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. *Psychological Science*, *31*(7), 770–780. https://doi.org/10.1177/0956797620939054

Pennycook, G., & Rand, D. G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(7), 2521–2526. https://doi.org/10.1073/pnas.1806781116

Pennycook, G., & Rand, D. G. (2021). The Psychology of Fake News. *Trends in Cognitive Sciences*, *25*(5), 388–402. https://doi.org/10.1016/j.tics.2021.02.007

Petty, R. E., Briñol, P., Tormala, Z. L., & Wegener, D. T. (2007). The role of metacognition in social judgment. In *Social psychology: Handbook of basic principles, 2nd ed.* (pp. 254–284). The Guilford Press.

Rabovsky, M., Stein, T., & Abdel Rahman, R. (2016). Access to awareness for faces during continuous flash suppression is not modulated by affective knowledge. *PLoS ONE*, *11*(4), e0150931. https://doi.org/10.1371/journal.pone.0150931

Rapp, D. N. (2016). The Consequences of Reading Inaccurate Information. *Current Directions in Psychological Science*, *25*(4), 281–285. https://doi.org/10.1177/0963721416649347

Rellecke, J., Sommer, W., & Schacht, A. (2012). Does processing of emotional facial expressions depend on intention? Time-resolved evidence from event-related brain potentials. *Biological Psychology*, *90*(1), 23–32. https://doi.org/10.1016/j.biopsycho.2012.02.002

Robbins, M. L., & Karan, A. (2020). Who Gossips and How in Everyday Life? *Social Psychological and Personality Science*, *11*(2), 185–195. https://doi.org/10.1177/1948550619837000

Rohr, L., & Abdel Rahman, R. (2015). Affective responses to emotional words are boosted in communicative situations. *NeuroImage*, *109*(C), 273–282. https://doi.org/10.1016/j.neuroimage.2015.01.031

Rohr, L., & Abdel Rahman, R. (2018). Loser! On the combined impact of emotional and person-descriptive word meanings in communicative situations. *Psychophysiology*, *55*(7), e13067--11. https://doi.org/10.1111/psyp.13067

Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, *5*(1), 1–10. https://doi.org/10.1057/s41599-019-0279-9

Sabatinelli, D., Keil, A., Frank, D. W., & Lang, P. J. (2013). Emotional perception: Correspondence of early and late event-related potentials with cortical and subcortical functional MRI. *Biological Psychology*, *92*(3), 513–519. https://doi.org/10.1016/j.biopsycho.2012.04.005

Schacht, A., & Sommer, W. (2009a). Time course and task dependence of emotion effects in word processing. *Cognitive, Affective and Behavioral Neuroscience*, *9*(1), 28–43. https://doi.org/10.3758/CABN.9.1.28

Schacht, A., & Sommer, W. (2009b). Emotions in word and face processing: Early and late cortical responses. *Brain and Cognition*, *69*(3), 538–550. https://doi.org/10.1016/j.bandc.2008.11.005

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Series in affective science.* (pp. 92–120). New York, NY, US: Oxford University Press.

Schindler, S., Bruchmann, M., Krasowski, C., Moeck, R., & Straube, T. (2021). Charged With a Crime: The Neuronal Signature of Processing Negatively Evaluated Faces Under Different Attentional Conditions. *Psychological Science*, *32*(8), 1311–1324. https://doi.org/10.1177/0956797621996667

Schindler, S., Bruchmann, M., Steinweg, A. L., Moeck, R., & Straube, T. (2020). Attentional conditions differentially affect early, intermediate and late neural responses to fearful and neutral faces. *Social Cognitive and Affective Neuroscience*, *15*(7), 765–774. https://doi.org/10.1093/scan/nsaa098

Schindler, S., & Bublatzky, F. (2020). Attention and emotion: An integrative review of emotional face processing as a function of attention. *Cortex*, *130*, 362–386. https://doi.org/10.1016/j.cortex.2020.06.010

Schindler, S., & Kissler, J. (2016). People matter: Perceived sender identity modulates cerebral processing of socio-emotional language feedback. *NeuroImage*, *134*(C), 160–169. https://doi.org/10.1016/j.neuroimage.2016.03.052

Schindler, S., Vormbrock, R., & Kissler, J. (2019). Emotion in context: How sender predictability and identity affect processing of words as imminent personality feedback. *Frontiers in Psychology*, *10*(FEB), 327. https://doi.org/10.3389/fpsyg.2019.00094

Schindler, S., Wegrzyn, M., Steppacher, I., & Kissler, J. (2015). Perceived communicative context and emotional content amplify visual word processing in the fusiform gyrus. *Journal of Neuroscience*, *35*(15), 6010–6019. https://doi.org/10.1523/JNEUROSCI.3346-14.2015

Schirmer, A. (2010). Mark my words: Tone of voice changes affective word representations in memory. *PLoS ONE*, *5*(2), e9080. https://doi.org/10.1371/journal.pone.0009080

Schöne, J. P., Parkinson, B., & Goldenberg, A. (2021). Negativity Spreads More than Positivity on Twitter After Both Positive and Negative Political Situations. *Affective Science*, *2*(4), 379–390. https://doi.org/10.1007/s42761-021-00057-7

Schumacher, P. B., Brandt, P., & Weiland-Breckle, H. (2018). Online Processing of "Real" and "Fake": The Cost of Being Too Strong. In *Language, Cognition, and Mind* (Vol. 4). https://doi.org/10.1007/978-3-319-77791-7_4

Schupp, H. T., Junghöfer, M., Öhman, A., Weike, A. I., Stockburger, J., & Hamm, A. O. (2004). The facilitated processing of threatening faces: An ERP analysis. *Emotion*, *4*(2), 189–200. https://doi.org/10.1037/1528-3542.4.2.189

Schupp, H. T., Junghöfer, M., Weike, A. I., & Hamm, A. O. (2003). Emotional facilitation of sensory processing in the visual cortex. *Psychological Science*, *14*(1), 7–13. https://doi.org/10.1111/1467-9280.01411

Schweinberger, S. R., & Neumann, M. F. (2016). Repetition effects in human ERPs to faces. *Cortex*, *80*, 141–153. https://doi.org/10.1016/j.cortex.2015.11.001

Simis, M. J., Madden, H., Cacciatore, M. A., & Yeo, S. K. (2016). The lure of rationality: Why does the deficit model persist in science communication? *Public Understanding of Science*, *25*(4), 400–414. https://doi.org/10.1177/0963662516629749

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. https://doi.org/10.1037/emo0000305

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*(3), 220–247. https://doi.org/10.1207/s15327957pspr0803_1

Suess, F., Rabovsky, M., & Abdel Rahman, R. (2013). Perceiving emotions in neutral faces: Expression processing is biased by affective person knowledge. *Social Cognitive and Affective Neuroscience*, *10*(4), 531–536. https://doi.org/10.1093/scan/nsu088

Suzuki, A., Honma, Y., & Suga, S. (2013). Indelible distrust: Memory bias toward cheaters revealed as high persistence against extinction. *Journal of Experimental Psychology: Learning Memory and Cognition*, *39*(6), 1901–1913. https://doi.org/10.1037/a0033335

Talarico, J. M., Berntsen, D., & Rubin, D. C. (2009). Positive emotions enhance recall of peripheral details. *Cognition and Emotion*, *23*(2), 380–398. https://doi.org/10.1080/02699930801993999

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

Todorov, A., & Olson, I. 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. *Social Cognitive and Affective Neuroscience*, *3*(3), 195–203. https://doi.org/10.1093/scan/nsn013

Tulling, M., Law, R., Cournane, A., & Pylkkänen, L. (2021). Neural correlates of modal displacement and discourse-updating under (Un)certainty. *ENeuro*, *8*(1), 1–19. https://doi.org/10.1523/ENEURO.0290-20.2020

Uhlmann, E. L., Pizarro, D. A., & Diermeier, D. (2015). A Person-Centered Approach to Moral Judgment. *Perspectives on Psychological Science*, *10*(1), 72–81. https://doi.org/10.1177/1745691614556679

van der Bles, A. M., van der Linden, S., Freeman, A. L. J., & Spiegelhalter, D. J. (2020). The effects of communicating uncertainty on public trust in facts and numbers. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(14), 7672–7683. https://doi.org/10.1073/pnas.1913678117

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

Wang, K., Goldenberg, A., Dorison, C. A., Miller, J. K., Uusberg, A., Lerner, J. S., Gross, J. J., Agesin, B. B., Bernardo, M., Campos, O., Eudave, L., Grzech, K., Ozery, D. H., Jackson, E. A., Garcia, E. O. L., Drexler, S. M., Jurković, A. P., Rana, K., Wilson, J. P., … Moshontz, H. (2021). A multi-country test of brief reappraisal interventions on emotions during the COVID-19 pandemic. *Nature Human Behaviour*, *5*(8), 1089–1110. https://doi.org/10.1038/s41562-021-01173-x

Wang, W. C., Brashier, N. M., Wing, E. A., Marsh, E. J., & Cabeza, R. (2016). On known unknowns: Fluency and the neural mechanisms of illusory truth. *Journal of Cognitive Neuroscience*, *28*(5), 739–746. https://doi.org/10.1162/jocn_a_00923

Webb, T. L., Miles, E., & Sheeran, P. (2012). Dealing with feeling: A meta-analysis of the effectiveness of strategies derived from the process model of emotion regulation. *Psychological Bulletin*, *138*(4), 775–808. https://doi.org/10.1037/a0027600

Wieser, M. J., & Brosch, T. (2012). Faces in context: A review and systematization of contextual influences on affective face processing. *Frontiers in Psychology*, *3*(NOV). https://doi.org/10.3389/fpsyg.2012.00471

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. https://doi.org/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. *Frontiers in Psychology*, *7*(AUG), 1446. https://doi.org/10.3389/fpsyg.2016.01258

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*(3), 441–517. https://doi.org/10.1006/jmla.2002.2864

Zillmann, D. (2013). Excitation Transfer Theory. In *Encyclopedia of Human Relationships* (Vol. 89). SAGE Publications, Inc. https://doi.org/10.4135/9781412958479.n181

Zimmerman, C. A., & Kelley, C. M. (2010). "I'll remember this!" Effects of emotionality on memory predictions versus memory performance. *Journal of Memory and Language*, *62*(3), 240–253. https://doi.org/10.1016/j.jml.2009.11.004

Danksagung

Von ganzem Herzen möchte ich mich zuerst bei meiner Doktormutter Rasha Abdel Rahman bedanken. Meine Dissertation und mein Aufwachsen in der Wissenschaft haben entscheidend von ihrer Förderung, ihrer positiven Einstellung, ihrem Scharfsinn, ihrer offenen und inspirierenden Begeisterung und Neugier für die neurokognitive Psychologie, ihrem großen Vertrauen in mich und meine Arbeit, ihrer konstanten, verlässlichen, schnellen und unkomplizierten Betreuung und nicht zuletzt ihrer Freundschaft profitiert und die Zeit der Promotion besonders kostbar gemacht.

Ich danke herzlich Sebastian Schindler und Reinhard Beyer für die Begutachtung meiner Dissertation, ihre anregenden Fragen und die Möglichkeit, meine Arbeit mit ihnen diskutieren zu können. Ich danke der gesamten Promotionskommission mit Martin Rolfs als Vorsitz und Ronja Demel als Mitglied, die die Verteidigung zu einer besonderen und unvergesslichen Erfahrung machen. Ich danke Jana Lahmer im Promotionsbüro für ihre Hilfsbereitschaft und Erreichbarkeit.

Mein großer Dank gehört dem Abdel Rahman Lab, allen Neuros, die durch die Kombination aus freundlicher Atmosphäre und Zusammenarbeit, anregendem Austausch und exzellenter Arbeit ein einzigartiges Umfeld geschaffen haben, von dem diese Dissertation profitieren durfte und was die Arbeit daran zum Vergnügen machte. Mein großer Dank geht an Guido Kiecker, ohne ihn, seine Erfahrung, sein Knowhow und seine technische Unterstützung diese Arbeit kaum möglich gewesen wäre. Für ihre Freundschaft, ihr Mentoring und ihre Hilfsbereitschaft danke ich allen Neuros, insbesondere danke ich dafür und für die schöne gemeinsame Zeit Kirsten Stark, Martin Maier, Alexander Enge, Anna Kuhlen, Milena Rabovsky und Romy Frömer. Ich danke allen studentischen Hilfskräften, mit denen ich zusammenarbeiten durfte und allen Studierenden, deren Abschlussarbeiten ich betreuen durfte, für ihre exzellente Arbeit, Unterstützung und Inspiration. Insbesondere danke ich Magdalena Kleissner, Martin Zielinski, Cornelius Braun, Josephine Dehnbostel und Niels Lilienthal. Ich danke Conny Reggentin für ihre immer ausgezeichnete und herzliche Unterstützung bei verschiedenen bürokratischen Hürden. Ich danke Chistine Kupper für ihre wegweisende Begrüßung bei den Neuros vor einigen Jahren. Ich danke unserem Kolloquium, insbesondere Werner Sommer, Olaf Dimigen und Peter Ullsperger für spannende Denkanstöße und wertvolles Feedback.

Ich danke Romy Frömer außerdem für ihr herzliches Willkommenheißen bei meinem Research Visit an der Brown University in Providence und für ihre exzellente Einführung in Single-Trial Analysen und Methoden.

Ich danke der Berlin School of Mind and Brain. Ich danke Isabel Dziobek für ihre Co-Betreuung, hilfreiche Treffen und die Möglichkeit in ihrem Kolloquium meine Arbeit zu diskutieren. Ich danke M&B für hilfreiche Kurse, Postersessions, die Aegina Summer School und Travel Grants, dabei danke ich insbesondere Annette Winkelmann und Dirk

Mende für ihre freundliche Unterstützung und Hilfsbereitschaft. Ich danke der M&B Kohorte 2017, allen Part-time Zombies, für unsere PhD Familie, unser Peer-Mentoring und unsere Freundschaft, insbesondere danke ich Cornelia van Scherpenberg, Luisa Balzus, Aroma Dabas, Sandra Naumann und Evelyn Medawar.

Ich danke der Studienstiftung des deutschen Volkes und dem Caroline von Humboldt Stipendiatinnenprogramm für Promotionsstipendien und ideelle Förderung und Rasha Abdel Rahman für Projektgelder für Experimente und meine weitere Finanzierung. Ich danke Astrid Dostert für ihre organisatorische Unterstützung beim Verlängern meiner Finanzierung.

Ich danke allen, die sich am Promovierendenaustausch am Institut beteiligen und beteiligt haben und besonders Richard Wundrack dafür, dass wir zum Beispiel den Stammtisch ins Leben gerufen haben. Ich danke Daniel Schad für seinen Workshop in Mixed Modeling für die Promovierenden am Institut und seine Hilfsbereitschaft bei unseren Fragen.

Von ganzem Herzen danke ich meiner Familie. Ich danke meinem Mann Pete, meinen Eltern Barbara und Jürgen und meinen Geschwistern Franziska und Alexander mit Familie für ihre Hilfe, ihre Geduld, ihre felsenfeste Unterstützung, ihr Vertrauen in mich, ihren Mut für mich und ihre Motivation. Ich danke meinen Freund*innen, dass auch sie immer da sind und jetzt diesen Meilenstein mit mir feiern.

Eigenständigkeitserklärung

Hiermit erkläre ich, die Dissertation selbstständig und nur unter Verwendung der angegebenen Hilfen und Hilfsmittel angefertigt zu haben.

Ich habe mich anderwärts nicht um einen Doktorgrad beworben und besitze keinen entsprechenden Doktorgrad.

Ich erkläre, dass ich die Dissertation oder Teile davon nicht bereits bei einer anderen wissenschaftlichen Einrichtung eingereicht habe und dass sie dort weder angenommen noch abgelehnt wurde.

Ich erkläre die Kenntnisnahme der dem Verfahren zugrunde liegenden Promotionsordnung der Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin vom 5. März 2015.

Weiterhin erkläre ich, dass keine Zusammenarbeit mit gewerblichen Promotionsbearbeiter*innen stattgefunden hat und dass die Grundsätze der Humboldt-Universität zu Berlin zur Sicherung guter wissenschaftlicher Praxis eingehalten wurden.

Julia Baum

Original Articles

1.

Baum, J., Rabovsky, M., Rose, S. B., & Abdel Rahman, R. (2018). Clear Judgments Based on Unclear Evidence: Person Evaluation Is Strongly Influenced by Untrustworthy Gossip. *Emotion*, *20*(2), 248–260. https://doi.org/10.1037/emo0000545

2.

Baum, J., & Abdel Rahman, R. (2020). Emotional news affects social judgments independent of perceived media credibility. *Social Cognitive and Affective Neuroscience*, *16*(3), 280–291. https://doi.org/10.1093/scan/nsaa164

3.

Baum, J., & Abdel Rahman, R. (2021). Negative news dominates fast and slow brain responses and social judgments even after source credibility evaluation. *NeuroImage*, *244*(August), 118572. https://doi.org/10.1016/j.neuroimage.2021.118572

Clear judgments based on unclear evidence:

Person evaluation is strongly influenced by untrustworthy gossip

Julia Baum*[1,2], Milena Rabovsky*[1,3], Sebastian Benjamin Rose[1] & Rasha Abdel Rahman[1,2]

[1] Humboldt-Universität zu Berlin

[2] Berlin School of Mind and Brain

[3] Freie Universität Berlin

Author Note

This work was supported by German Research Foundation grant AB 277-6 to Rasha Abdel Rahman

Corresponding Authors:

Julia Baum, julia.baum@hu-berlin.de, Rudower Chaussee 18, 12489 Berlin

Rasha Abdel Rahman, rasha.abdel.rahman@hu-berlin.de, Rudower Chaussee 18, 12489 Berlin

* indicates equal contribution

Abstract

Affective information about other people's social behavior may prejudice social interactions and bias person judgments. The trustworthiness of person-related information, however, can vary considerably, as in the case of gossip, rumours, lies, or so-called "fake news". Here, we investigated how spontaneous person-likability and explicit person judgments are influenced by trustworthiness, employing event-related potentials as indexes of emotional brain responses. Social-emotional information about the (im)moral behaviour of previously unknown persons was verbally presented as trustworthy fact, (e.g. "*He bullied his apprentice"*) or marked as untrustworthy gossip (by adding e.g. *allegedly*), using verbal qualifiers that are frequently used in conversations, news and social media to indicate the questionable trustworthiness of the information and as a precaution against wrong accusations. In Experiment 1, spontaneous likability, deliberate person judgments and electrophysiological measures of emotional person evaluation were strongly influenced by negative information, yet remarkably unaffected by the trustworthiness of the information. Experiment 2 replicated these findings and extended them to positive information. Our findings demonstrate a tendency for strong emotional evaluations and person judgments even when they are knowingly based on unclear evidence.

Keywords: trustworthiness, gossip, face perception, person evaluation, event-related potentials

Clear judgments based on unclear evidence:

Person evaluation is strongly influenced by untrustworthy gossip

Praised be doubt! I advise you to greet
Cheerfully and with respect the man
Who tests your word like a bad penny!
(Bertholt Brecht, 1932)

The veracity of person-related verbal information is often questionable. This has lately been a prominent observation in public communication, where people encounter gossip, rumours, and news from sources of varying reliability on a daily basis. As yet, little is known about how verbally communicated information of questionable reliability affects us. The goal of the present study was to investigate how one considers the lack of reliability when one subjectively judges persons as negative, neutral or positive based on verbal information about their immoral or moral social behavior.

How people perceive, judge, and interact with others is strongly influenced by what they know about them. Even abstract and verbally transmitted information concerning good or bad social behavior can affect how they judge others (Bliss-Moreau, Barrett, & Wright, 2008; Goodwin, Piazza, & Rozin, 2014), how they perceive others' faces or facial expressions (Abdel Rahman, 2011; Luo, Wang, Dzhelyova, Huang, & Mo, 2016; Suess, Rabovsky, & Abdel Rahman, 2015; Wieser et al., 2014; Xu, Li, Diao, Fan, & Yang, 2016), and may even affect whether they see others' faces in the first place (Anderson, Siegel, Bliss-Moreau, & Barrett, 2011; but see Rabovsky, Stein, & Abdel Rahman, 2016; Stein, Grubb, Bertrand, Suh, & Verosky, 2017). Here we consider one factor that may influence the potency of social-emotional information to modulate person evaluations: the verbally marked trustworthiness of the information. We did so by adding qualifiers like "allegedly", "supposedly", or "people say" to

person-related information, expressions often encountered during gossip-laden conversations and typically used to express doubt concerning the veracity of the information, for example in legal or journalistic contexts.

Via associations with affective person knowledge, faces gain intrinsic emotional relevance, leading to motivated attention at perceptual and post-perceptual evaluative processing stages (e.g., Abdel Rahman, 2011; Sabatinelli, Keil, Frank, & Lang, 2013). In event-related brain potentials (ERPs) derived from the EEG[1], high-level evaluation is reflected in an enhanced late positivity at about 400 to 600 ms over centro-parietal regions (late positive potential [LPP]; Schupp, Junghöfer, Weike, & Hamm, 2003; Schacht & Sommer, 2009). At earlier stages, affective information may induce an enhanced early posterior negativity (EPN) at about 200 to 300 ms at occipito-temporal sites related to fast and reflexive perception-related processes (Schupp et al., 2003).

According to appraisal theories of emotion (Ellsworth & Scherer, 2003; Scherer, 2001), stimuli are initially checked for a coarse detection of emotional salience and intrinsic pleasantness, a process that may be related to early ERP modulations as the EPN (e.g., Herbert, Pauli, & Herbert, 2011). Then, assessments concerning implications for the observer's well-being, coping possibilities, and evaluations of the normative significance – like the compatibility with moral standards – follow that can be related to higher level evaluations associated with LPP-generating processes (e.g., Herbert et al., 2011; Yoder & Decety, 2014). Crucially, whereas the LPP has been shown to vary with the relevance and meaning of emotional attributes in a

[1] We used ERPs because they offer information on processes and their modulation by experimental factors that cannot be directly observed. For instance, here, we expect a dissociation between effects in early (EPN) and later (LPP) components associated with distinct processes.

given context, the earlier emotion-induced reflexive EPN modulations are relatively independent of task demands and the relevance of emotional contents and are mainly sensitive to arousal (e.g., Hinojosa, Méndez-Bértolo, & Pozo, 2010; Kissler, Herbert, Winkler, & Junghöfer, 2009; Rellecke, Sommer, & Schacht, 2012; Schacht & Sommer, 2009). Specifically, the emotional content associated with a face may be appraised independently of the verbally marked trustworthiness of the information at early stages reflected in the EPN, whereas later stages of high-level evaluations reflected in the LPP – in which emotion effects are more strongly affected by context and relevance – should be more sensitive to additional information putting emotional contents into perspective and should therefore be modulated by the verbally marked trustworthiness of the information.

Empirical evidence on the trustworthiness of verbally transmitted information and its effects on person judgments is scarce. However, related research has provided evidence that emotional responses and person evaluation can be modulated by intentional emotion regulation and context information. Indeed, we can in many ways deliberately choose to ignore information (Hertwig & Engel, 2016). In this sense, untrustworthy person-related information may be deliberately ignored to achieve fair, unbiased social judgments. In line with these assumptions, influences of person knowledge on the spontaneous likability of faces were found to be reduced when participants were instructed to suppress affective verbal information previously associated with the faces (Molet et al., 2016), with stronger suppression for prosocial (e.g., “threw a surprise party for a parent”) compared to antisocial (e.g., “hit a small child”) information. Furthermore, changing the meaning of an emotional stimulus via context information, for example, by labeling an unpleasant scene as fictitious versus real (Mocaiber et al., 2010), or by reappraising a person’s angry face with the person’s bad day at work (Blechert, Sheppes, Di

Tella, Williams, & Gross, 2012) induces spontaneous emotion regulation reflected in attenuated LPP amplitudes (Foti & Hajcak, 2008). These studies suggest that people use context information to adjust their emotional reactions. On the other hand, context may also be involuntarily ignored, such as mistakenly associating social-emotional information with the wrong person, even though the correct context information is clearly available (Ecker, Lewandowsky, Chang, & Pillai, 2014). Specifically in the case of negative information, however, context information about the trustworthiness may be ignored reflexively or deliberately as a protection against potential threat (e.g., Öhman & Mineka, 2001).

The present study investigated effects of negative and positive person-related information that was associated with the faces of previously unfamiliar persons either as trustworthy facts (e.g., "He bullied his apprentice"; "He rescued refugees") or as untrustworthy gossip (e.g., "He is believed to have bullied his apprentice"; "He supposedly rescued refugees"). Gossip was explicitly verbally labeled as untrustworthy, enabling participants to doubt the information. Emotional information was compared to a neutral condition (e.g., "He visited clients"). Negative information was about harmful social behaviour, that is, immoral, and positive information was about kind social behaviour, that is, moral (*cf.* Hofmann, Wisneski, Brandt, & Skitka, 2014). Subsequent to learning person-related information participants completed a test phase in which they were instructed to make explicit person judgments based on the information they had learned, with the intention to motivate evaluative processing and consideration of the information's trustworthiness. Additionally, participants rated the person likability before and after learning, which served two purposes. First, it allowed us to compare a judgment in which the person-related information is directly task-relevant to a judgment in which this information is more indirectly relevant. Second, because this rating can be performed spontaneously without

any additional information, a comparison between the likability before and after learning is possible.

Based on the theoretical approaches and the empirical evidence discussed above, we expected effects of emotional relative to neutral information on explicit person evaluations and on EPN and LPP amplitudes, as documented in several studies (see above). Crucially, verbally marked trustworthiness should modulate person judgments, resulting in reduced or absent effects of untrustworthy compared to trustworthy emotional information. Early reflexive processing should be immune to contextual factors or task demands (see discussion above). Therefore, we expected EPN modulations of similar magnitude for faces associated with trustworthy and untrustworthy emotional relative to neutral information. In contrast, later evaluative processing should be susceptible to contexts and the (task) relevance of emotion. This should be reflected in reduced LPP amplitudes for untrustworthy relative to trustworthy emotional information. These effects may be modulated by valence, such that negative, but not positive, gossip may be prioritized even when verbally marked as untrustworthy.

Experiment 1

Method

Sample size

For experiment 1 a multiple of eight participants was required to counterbalance the assignment of information to faces. G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007) estimated a sample of 39 to 30 participants (alpha = .05, power = .9, three measurement levels, within subject rmANOVA) for a η_p^2 reported between .15 (Abdel Rahman, 2011), and .17 (Luo et

al., 2016) for LPP effects, and between .18 (Luo et al., 2016), and .19 (Suess et al., 2015) for EPN effects.

Participants

Data consist of 32 German participants (25 female, mean age = 25, *SD* = 4.98, 30 right-handed) with normal or corrected-to-normal vision. One participant was replaced due to insufficient learning (recollected less than 50%). Participants received course credit or were monetarily compensated. They were (de)briefed about the procedures and signed informed consent. The study was approved by the local ethics committee.

Materials

Picture stimuli were grey-scaled frontal portrait photographs of 64 unfamiliar faces with neutral facial expressions (2.7 × 3.5 cm, viewing distance 70 cm; obtained from various databases, e.g. Ebner et al., 2010; Langner et al., 2010). Sixteen familiar filler faces (well-known persons, e.g. Romano Prodi (neutral), or Josef Fritzl (negative)) were included to make the target persons' existence credible.

Short sentences describing social behavior were recorded (see Supplemental Material, Table S1). Information was either neutral (e.g. "She showed the new collection to a customer"), negative and presented as trustworthy fact (e.g. "She drowned her baby in the bathtub"), or negative and presented as untrustworthy gossip, (e.g. "People say she drowned her baby in the bathtub"). Neutral information was not presented as untrustworthy. The reason to not include an untrustworthy neutral condition was simply that this would sound ironic or irritating (e.g. "He allegedly consulted a technician") and may therefore have reduced the credibility of the information in general. Gossip was verbally marked as untrustworthy by using, for example, *people assume*, *allegedly*, *supposedly*, or *is believed to*. Thirty-two faces were assigned to neutral

information, 16 to negative facts, and 16 to negative gossip, with counterbalanced assignment across participants. Affective information for familiar filler faces referred to their biography and was not counterbalanced (8 neutral, 4 negative facts, 4 negative gossip; note that gossip referred to ongoing speculations or accusations, e.g. “He allegedly stabbed his wife’s lover” (OJ Simpson); “This man supposedly sexually harassed underage girls” (Silvio Berlusconi), whereas facts referred to convictions, e.g. “He committed a massacre to teenagers at a summer camp” (Anders Breivik)).

Manipulation check: Trustworthiness rating

After the experiment, we conducted an online rating to test whether the verbal marking was sufficient to reveal the trustworthiness differences. A sample size of the multiple of two participants was required for counterbalancing. G*Power estimated 36 participants for a medium effect size of dz = .5 (alpha = .05, power = .9, one-tailed paired t-test). A different group of participants (N = 38, mean age = 28.20, SD = 7.05, 21 female) was instructed to rate how trustworthy they consider each individually presented information about unfamiliar persons. The 11-point scale ranged from 1 (*not at all trustworthy*) to 11 (*very trustworthy*). Only negative information was included, with the presentation as fact or gossip counterbalanced across raters. Gossip was rated as less trustworthy than facts, $t(37) = 3.42$, $p = .002$, dz = .56, (see Table 1). This indicates that the trustworthiness manipulation served the intended purpose and that, accordingly, our experimental manipulation was successful.

Table 1

Manipulation Check in Experiment 1: Trustworthiness Ratings of the Person-Related Information. Ratings on an 11-Point Scale from not at all trustworthy (1) to very trustworthy (11)

Information	Negative Facts	Negative Gossip
M	6.65	5.70
95% CI	[6.25, 7.05]	[5.29, 6.10]

Note. CI = confidence interval.

Procedure

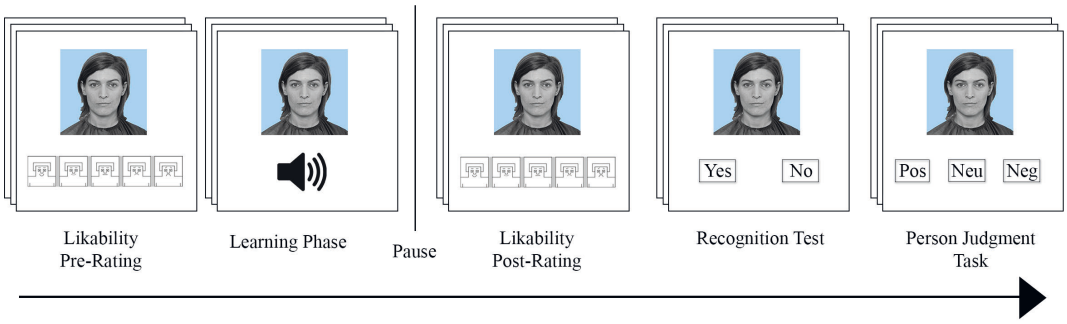

Figure 1. Schematic illustration of the experimental procedure. Event-related potential effects were investigated during the person judgment task. *Consent to publish photo here has been given by depicted person.*

To create the global impression that all people encountered during the experiment existed in real life, we included well-known filler faces, and participants were told that they would be seeing faces and that some of them they may be more familiar with than others. The experiment had a learning and a testing phase (*cf.* Figure 1), separated by a 15-minute break. In the beginning of each phase, participants rated how likable they spontaneously found each person on a 5-point scale (adopted from the Self-Assessment Manikin; Bradley & Lang, 1994). We call this rating "spontaneous likability", because this rating is not based on the information, but rather on the spontaneous feeling of liking; that is, the information was not task relevant. During learning,

participants saw the face and listened to the assigned verbal information. Block-wise learning familiarized the participants with 10 faces at a time. Each block included all experimental conditions and two filler faces, with the exact combination of stimuli being random across participants (for a similar design, see Abdel Rahman, 2011; Suess et al., 2015). Across learning, faces were presented six times with information (for the duration of the recorded sentence, on average 3.4 s, see Supplemental Material) and one last time without. To ensure participants paid attention, they answered yes-or-no questions about the information learned in a block, e.g. “Is the behavior of this person common?” (asked per face, four questions in total).

The test phase always started with the likability rating (see learning phase). A recognition test and the person judgment task followed, in a counterbalanced sequence. To test recognition, participants decided whether a face was familiar from the learning phase (included 80 unfamiliar filler faces). For the person judgment, participants had to judge whether the depicted person was negative or neutral to them based on the information acquired during learning. Answer button assignment was counterbalanced.

Each task was repeated six times, resulting in 192 person judgment trials for faces with neutral information, and 192 for negative, i.e. 96 negative facts, 96 negative gossip. Trials started with a 500 ms pre-stimulus fixation cross, faces were presented until response or for 2 s. In the likability rating, faces were presented for 1 s, followed by the rating scale until response.

After the experiment, a paper-and-pencil test checked for sufficient learning. The experiment continued on a second test day that is not part of the present research question and not discussed further, as it does not affect the first day. Due to this second test day, counterbalancing required a sample size of a multiple of eight participants. The recognition test

was planned to control for the learning of the faces and mainly relevant for the second day. Here, we focused on the likability rating and person judgment.

Data recording and analysis

The EEG was recorded from 62 electrode sites as specified by the extended 10-20 system with Ag/AgCl electrodes. Impedance was kept under 5 kΩ. The sampling rate was 500 Hz, and the continuous signal was referenced to the left mastoid. 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, later used to correct for eye movement artifacts.

Offline, the continuous EEG was transformed to average reference and low-pass filtered at 30 Hz (24dB/oct, zero-phase IIR Butterworth filter). Using BESA (Berg & Scherg, 1991), we removed artifacts due to eye movements by applying a spatiotemporal dipole modeling procedure for each participant individually. A semi-automatic procedure rejected remaining artifacts by filtering out amplitudes over ± 200 μV, changing more than 50 μV between samples or more than 200 μV within single epochs, or containing baseline drifts. Error- and artifact-free EEG data was segmented into epochs of 2.5 s, starting 100 ms prior to stimulus onset (i.e. appearance of a face during the judgment task), with a 100 ms pre-stimulus baseline.

We performed repeated measures ANOVAs (rmANOVAs) on averaged data per information condition (neutral vs. negative facts vs. negative gossip) to assess amplitude differences in ERPs during the person judgment task. Because trustworthiness was not manipulated in the neutral condition (lack of credibility in the neutral condition), we calculated a main effect over all three conditions and followed up with separate comparisons. Analyses focused on two regions of interest, based on previous findings of emotional stimulus content

(e.g. Schupp et al., 2003) and affective information effects (e. g. Abdel Rahman, 2011; Suess et al., 2015) in the EPN (electrode sites PO7, PO8, PO9, PO10, TP9, TP10, between 200 and 300 ms) and LPP component (electrode sites Pz, CPz, POz, P3, P4, between 400 and 600 ms). Huyhn-Feldt corrections were applied when appropriate. We report uncorrected degrees of freedom and in case of separate comparisons Bonferroni corrected p-values for the number of analyses. The significance level was $p < .05$. Data and Code are available from the Open Science Framework (osf.io/jqv2g; Baum, Rabovsky, Rose, & Abdel Rahman, 2018).

Results

Behavioral results

For results of the recognition test see Supplemental Material (Table S2).

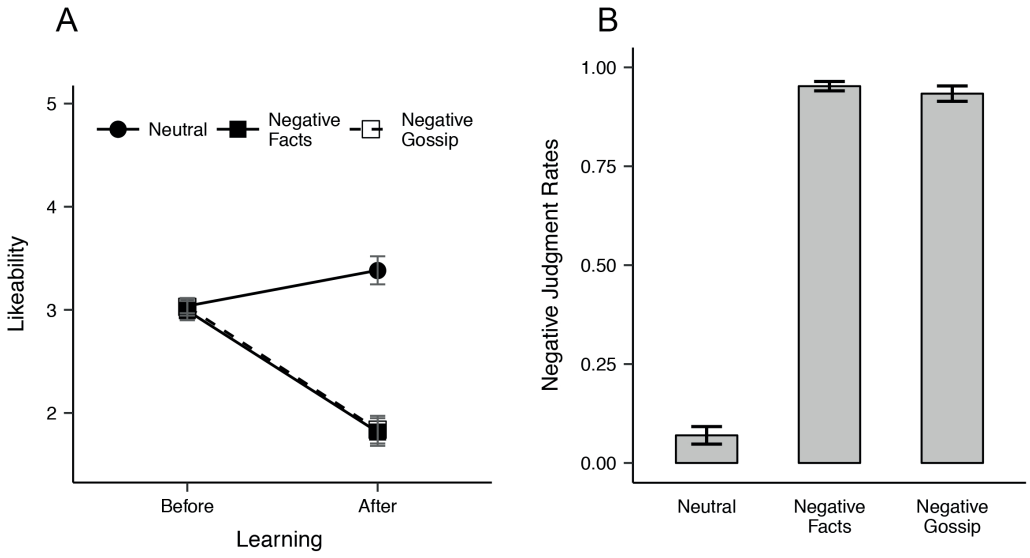

Figure 2. Behavioral results of experiment 1. Error bars show 95% confidence intervals. (A) Mean likability ratings of persons before and after learning person-related information, depending on type of information. (B) Mean judgment rates in the person judgment task, depending on type of information.

Likability rating

Whereas there was no difference in likability before learning, trustworthy as well as untrustworthy negative information led participants to dislike persons relatively to persons associated with neutral information (see Figure 2a and Table 2). A rmANOVA revealed main effects of phase (two levels: before vs. after learning), $F(1,31) = 118.57$, $p < .001$, $\eta_p^2 = .79$, $\eta_G^2 = .56$, and information (three levels: neutral, negative facts, negative gossip), $F(2,62) = 173.44$, $p < .001$, $\eta_p^2 = .85$, $\eta_G^2 = .60$, as well as an interaction of phase and information, $F(2,62) = 110.68$, $p < .001$, $\eta_p^2 = .78$, $\eta_G^2 = .59$. Information had no effect before learning, $F(2,62) = .58$, $p = 1$, $\eta_G^2 = .018$, but after learning, $F(2,62) = 163.23$, $p < .001$, $\eta_G^2 = .84$. Faces associated with negative facts and negative gossip were rated as less likable than faces with neutral information; $F(1,31) = 169.37$, $p < .001$, $\eta_G^2 = .85$, and $F(1,31) = 172.24$, $p < .001$, $\eta_G^2 = .85$ respectively. Facts and gossip did not differ, $F(1,31) = .41$, $p = 1$, $\eta_G^2 = .013$.

Person judgment

Persons were judged as negative based on negative facts and negative gossip compared to the neutral condition (see Figure 2b and Table 3). Judgment rates differed by information, $F(2,62) = 2398.86$, $p < .001$, $\eta_G^2 = .99$. Compared to neutral information, faces associated with negative facts were more frequently judged as negative, $F(1,31) = 4197.12$, $p < .001$, $\eta_G^2 = .99$, and likewise, faces associated with negative gossip were more frequently judged as negative, $F(1,31) = 2226.15$, $p < .001$, $\eta_G^2 = .99$. Judgments based on negative facts and negative gossip did not differ, $F(1,31) = 3.10$, $p = .27$, $\eta_G^2 = .09$.

Faces associated with both negative facts and negative gossip were judged faster than faces with neutral information, with no difference in reaction times (RTs) for facts and gossip (see Table 4 and 5).

Table 2

Likability rating of persons before and after learning person-related information in experiment 1 on a 5-point scale

Phase	Information	Neutral	Negative Facts	Negative Gossip
Before	*M*	3.04	2.99	3.03
	95% CI	[2.96, 3.11]	[2.89, 3.09]	[2.94, 3.12]
After	*M*	3.38	1.82	1.84
	95% CI	[3.24, 3.53]	[1.67, 1.96]	[1.69, 1.98]

Note. CI = confidence interval.

Table 3

Means and confidence intervals for negative person judgments in experiment 1

Information	Neutral	Negative Facts	Negative Gossip
M	.070	.95	.93
95% CI	[.044, .095]	[.94, .97]	[.91, .96]

Note. CI = confidence interval.

Table 4

Means and confidence intervals for reaction times (in ms) for person judgments in experiment 1

Information	Neutral	Negative Facts	Negative Gossip
M	831.35	777.43	784.51
95% CI	[817.33, 845.37]	[762.19, 792.67]	[772.19, 796.82]

Note. CI = confidence interval.

Table 5

Statistical results for reaction times for person judgments in experiment 1

Source	*df*	*F*	*p*	η_G^2
Negative Facts *vs*. Negative Gossip *vs.* Neutral	2,62	18.46	< .001	.37
Negative Facts *vs*. Neutral	1,31	25.70	< .001	.45
Negative Gossip *vs.* Neutral	1,31	29.48	< .001	.49
Negative Facts *vs*. Negative Gossip	1,31	.055	1	.017

Event-Related Potentials

EPN. No main effect of information was found, $F(2,62) = .70$, $p = .50$, $\eta_p^2 = .022$, $\eta_G^2 < .001$.

LPP. Negative facts as well as negative gossip elicited an enhanced positivity compared to neutral information (see Figure 3), reflected in a main effect of information, $F(2,62) = 8.57$, $p = .001$, $\eta_p^2 = .22$, $\eta_G^2 = .036$. Separate analyses revealed an enhanced positivity for negative facts compared to neutral information, $F(1,31) = 7.64$, $p = .03$, $\eta_p^2 = .20$, $\eta_G^2 = .033$, as well as for negative gossip compared to neutral information, $F(1,31) = 13.54$, $p = .003$, $\eta_p^2 = .30$, $\eta_G^2 = .046$. Faces associated with facts and gossip did not differ, $F(1,31) = .60$, $p = 1$, $\eta_p^2 = .02$, $\eta_G^2 = .001$.

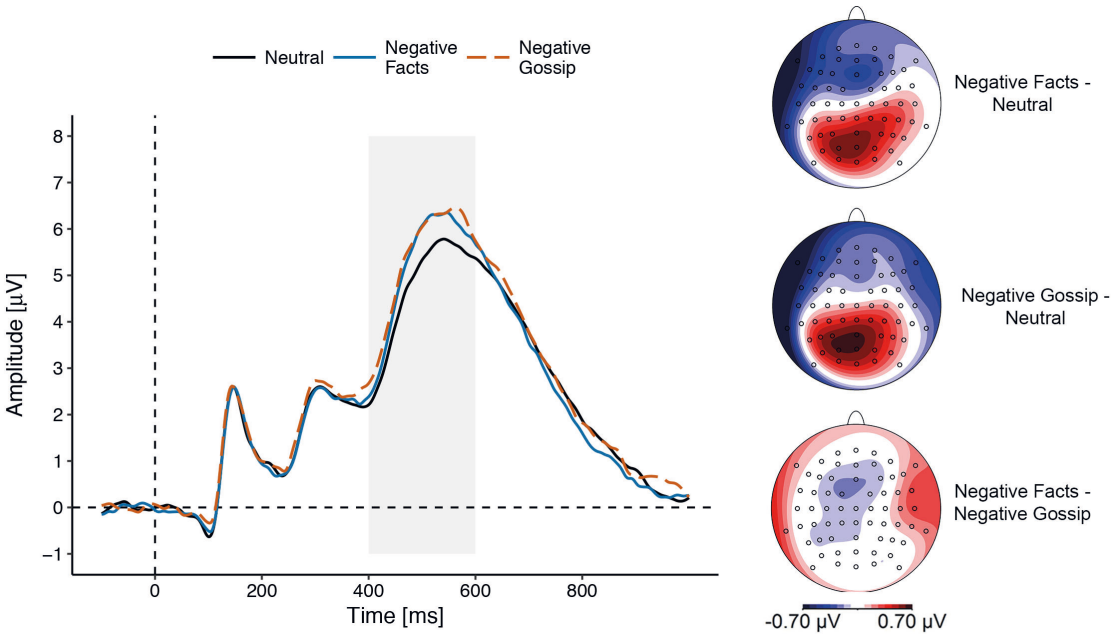

Figure 3. Grand average event-related potentials at the central-posterior site Pz show the LPP effects of person-related information in the person judgment task in experiment 1. Time point zero marks onset of face presentation. Scalp distributions show the effects as differences between conditions in the LPP time window 400 to 600 ms (grey area).

Discussion

We expected effects of emotional relative to neutral information on EPN and LPP amplitudes and explicit person judgments. Verbally marked trustworthiness should result in reduced emotion effects of untrustworthy compared to trustworthy information, reflected in the LPP and explicit judgments. We expected EPN amplitudes to be unaffected by trustworthiness.

Experiment 1 showed no effects of information on the EPN, where we had expected emotion effects. EPN emotion effects are very robust for well-known faces, however more vulnerable if the faces have been newly learned (Abdel Rahman, 2011; Suess et al., 2015). As expected, we found that negative person-related information strongly affected spontaneous

ratings of likability, explicit person judgments and LPP amplitudes reflecting evaluative processes. Additionally, the reaction times in the person judgment show that the negative judgments based on gossip were made without hesitation. Unexpectedly, however, none of these effects was modulated by the trustworthiness of the information, even though gossip could be identified as untrustworthy, as shown by an independent manipulation check.

Experiment 2

Experiment 2 was conducted to replicate the results of experiment 1 and to investigate if the findings generalize to effects of positive information. Positive information was included to test the possibility that the trustworthiness information was involuntarily or deliberately ignored to cope with the potential threat that is related to the negatively valenced person information. Thus, we tested the hypothesis that the untrustworthiness of non-threatening, positive gossip may be considered, resulting in reduced emotion effects of untrustworthy compared to trustworthy positive information, reflected in the LPP and explicit judgments. We expected the emotion effects in the EPN to be unaffected by trustworthiness.

Methods

Methods of experiment 2 were identical to experiment 1, except for the details described in the following.

Sample size

A multiple of 6 participants was required for a counterbalanced assignment of information conditions to faces. For an effect size of $\eta_p^2 = .22$ (main effect of information in the LPP in experiment 1), G*Power estimated a sample of 25 participants (alpha .05, power .9, number of measurements 3).

Participants

Data consists of 24 German participants (16 female, mean age = 25.17 years (*SD* = 5.76), range 18 to 44, all right-handed). One was replaced due to not following task instructions, and two due to artefacts in the EEG.

Materials

Picture stimuli were 24 unfamiliar faces and 8 familiar filler faces; we reduced the total number of stimuli to facilitate learning. Social information's valence was neutral, negative, or positive. The selection of sentences was based on an independent rating, making sure both affective conditions were considerably more arousing than neutral information (no trustworthiness manipulation yet; see Supplemental Material). Negative and positive information was either presented fact-like (e.g., positive fact: *She resuscitated a tourist)*, or gossip-like (e.g., positive gossip: *She is believed to have resuscitated a tourist*; see Supplemental Material, Table S3). Eight unfamiliar faces were assigned to neutral information, 4 to negative facts, 4 to positive facts, 4 to negative gossip, and 4 to positive gossip, counterbalanced across participants. Affective information for familiar filler faces referred to their biography (4 neutral, 2 negative facts, 2 positive facts).

Procedure

During learning, one block consisted of 8 faces (1 negative fact, 1 negative gossip, 1 positive fact, 1 positive gossip, 2 neutral, 1 neutral filler and 1 negative or positive filler). Across learning, faces were presented 4 times with information and one last time without, and 3 learning enhancing questions were answered per face.

The test phase consisted of the likability rating, a recognition test, and the person judgment task in this order. This was because the rating and the recognition test were performed

only once, and the focus was now on the person judgment task, which was repeated 20 times, resulting in 160 trials for faces with neutral information, 160 for negative (80 facts, 80 gossip), and 160 for positive (80 facts, 80 gossip).

After the experiment, a trustworthiness rating of the learned information was conducted with the same participants. Each sentence was visually presented on screen as it had been learned but without the assigned face. Participants were instructed to rate how trustworthy they considered the information about the person's behavior on an 11-point scale. The neutral information was presented as fillers, since it did not contain a trustworthiness manipulation.

Data recording and analysis

Analogously to experiment 1, we analysed effects (neutral vs. facts vs. gossip) for negative and positive conditions separately. Additionally, we performed rmANOVAs including the factors valence (negative vs. positive) and trustworthiness (facts vs. gossip) to investigate possibly different effects of trustworthiness depending on the valence of the information. Because trustworthiness was not manipulated in the neutral condition, it was not included in those analyses.

Results

Behavioral results

For results of the recognition test see Supplemental Material (Table S4).

Trustworthiness rating

Facts were rated more trustworthy than gossip (see Table 6). A rmANOVA with the factors valence (negative vs. positive) and trustworthiness (facts vs. gossip) revealed a main effect of trustworthiness, $F(1,23) = 6.84$, $p = .015$, $\eta_p^2 = .23$, $\eta_G^2 = .13$, but no effect of valence,

$F(1,23) = 0.46$, $p = .51$, $\eta_p^2 = .02$, $\eta_G^2 = .005$, and no interaction of valence and trustworthiness, $F(1,23) = 1.08$, $p = .31$, $\eta_p^2 = .04$, $\eta_G^2 = .01$.

Table 6

Manipulation Check in experiment 2: Trustworthiness ratings of the verbal person-related information. Ratings on an 11-point scale from not at all trustworthy (1) to very trustworthy (11).

Information	Negative facts	Negative gossip	Positive facts	Positive gossip
M	7.70	6.45	7.63	6.90
95% CI	[6.98, 8.42]	[5.95, 6.95]	[7.00, 8.25]	[6.22, 7.57]

Note. CI = confidence interval.

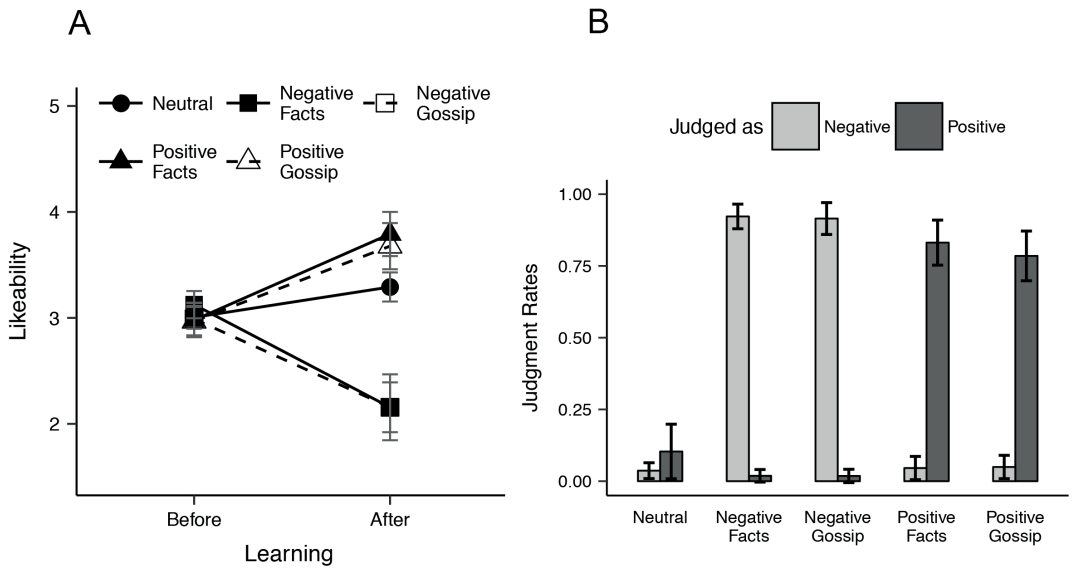

Figure 4. Behavioral results of experiment 2. Error bars show 95% confidence intervals. (A) Mean likability ratings of persons before and after learning person-related information depending on type of information. (B) Mean judgment rates in the person judgment task depending on type of information.

Likability rating

Facts and gossip led participants to like or dislike persons more relatively to the neutral condition, depending only on the valence of the information (see Figure 4a and Table 7).

In the negative condition an analysis including experimental phase (2 levels: before vs. after learning) and information (negative facts, negative gossip, neutral) revealed a main effect of phase, $F(1,23) = 27.21$, $p < .001$, $\eta_p^2 = .54$, $\eta_G^2 = .26$, of information, $F(2,46) = 20.27$, $p < .001$, $\eta_p^2 = .47$, $\eta_G^2 = .27$, and an interaction of phase and information, $F(2,46) = 36.51$, $p < .001$, $\eta_p^2 = .61$, $\eta_G^2 = .31$. Likability did not differ before, $F(2,46) = 1.29$, $p = .57$, $\eta_G^2 = .053$, but after learning, $F(2,46) = 35.75$, $p < .001$, $\eta_G^2 = .61$. Faces associated with negative facts and negative gossip were later rated less likable than faces associated with neutral information, $F(1,23) = 36.62$, $p < .001$, $\eta_G^2 = .61$ and $F(1,23) = 53.22$, $p < .001$, $\eta_G^2 = .70$ respectively. Facts and gossip did not differ, $F(1,23) = 0$, $p = 1$, $\eta_G^2 = 0$.

In the positive condition a main effect of phase, $F(1,23) = 62.78$, $p < .001$, $\eta_p^2 = .73$, $\eta_G^2 = .43$, information, $F(2,46) = 4.55$, $p = .016$, $\eta_p^2 = .17$, $\eta_G^2 = .078$, and an interaction of phase and information, $F(2,46) = 8.73$, $p = .001$, $\eta_p^2 = .28$, $\eta_G^2 = .10$, were found. There were no condition differences before, $F(2,46) = .071$, $p = 1$, $\eta_G^2 = .003$, but after learning, $F(2,46) = 11.22$, $p < .001$, $\eta_G^2 = .33$. Faces associated with positive facts and positive gossip were rated more likable than faces associated with neutral information, $F(1,23) = 23.74$, $p < .001$, $\eta_G^2 = .51$ and $F(1,23) = 13.26$, $p = .004$, $\eta_G^2 = .37$ respectively. Facts and gossip did not differ, $F(1,23) = 0.88$, $p = 1$, $\eta_G^2 = .037$.

An analysis after learning including the factors valence (negative, positive) and trustworthiness (facts, gossip), excluding the neutral condition, revealed a main effect of valence, $F(1, 23) = 68.44$, $p < .001$, $\eta_p^2 = .75$, $\eta_G^2 = .68$, no effect of trustworthiness, $F(1, 23) = .50$, p

= .49, η_p^2 = .021, η_G^2 = .003, and no interaction of valence and trustworthiness, $F(1, 23) = .46$, p = .51, η_p^2 = .019, η_G^2 = .003.

Person judgment

Compared to the neutral condition, judgments were more negative in the negative facts and in the negative gossip condition and more positive in the positive facts and in the positive gossip condition (see Figure 4b and Table 8).

In the negative condition there was a main effect of information, $F(2,46) = 730.32$, p < .001, η_G^2 = .97. Relative to faces connected to neutral information, faces connected to negative facts, $F(1,23) = 1225,53$, p < .001, η_G^2 = .98, and also faces connected to negative gossip, $F(1,23) = 1057.32$, p < .001, η_G^2 = .98, were more frequently judged as negative. Judgments did not differ for facts and gossip, $F(1,23) = .070$, p = 1, η_G^2 = .003.

In the positive condition there was an effect of information, $F(2,46) = 84.53$, p < .001, η_G^2 = .79. Relative to faces connected to neutral information, faces connected to positive facts, $F(1,23) = 121.55$, p < .001, η_G^2 = .84, and also faces connected to positive gossip, $F(1,23)$ = 104.56, p < .001, η_G^2 = .82, were more frequently judged as positive. Judgments did not differ for facts and positive gossip, $F(1,23) = .72$, p = 1, η_G^2 = .030.

Excluding the neutral condition, we found a main effect of valence, $F(1,23) = 15.10$, p < .001, η_p^2 = .40, η_G^2 = .17, but no effect of trustworthiness, $F(1,23) = .73$, p = .40, η_p^2 = .031, η_G^2 = .012, and no interaction of valence and trustworthiness, $F(1,23) = .43$, p = .52, η_p^2 = .018, η_G^2 = .006.

Table 7

Likability rating of persons before and after learning person-related information in experiment 2 on a 5-point scale

Phase	Info.	Neutral	Negative Facts	Negative Gossip	Positive Facts	Positive Gossip
Before	*M*	3.01	3.13	2.99	2.98	2.97
	95% CI	[2.89, 3.12]	[2.99, 3.26]	[2.83, 3.15]	[2.81, 3.15]	[2.83, 3.11]
After	*M*	3.29	2.16	2.16	3.79	3.68
	95% CI	[3.15, 3.44]	[1.83, 2.48]	[1.91, 2.40]	[3.57, 4.01]	[3.45, 3.90]

Note. CI = confidence interval.

Table 8

Means and confidence intervals for negative person judgments (row 1) and for positive person judgments (row 2) in experiment 2

Judgment	Info.	Neutral	Negative Facts	Negative Gossip	Positive Facts	Positive Gossip
Negative	*M*	.037	.92	.92	.046	.049
	95% CI	[.009 .064]	[.88, .97]	[.86, .97]	[.005, .086]	[.009, .090]
Positive	*M*	.10	.019	.018	.83	.78
	95% CI	[.008, .19]	[-.003, .041]	[-.005, .042]	[.75, .91]	[.70, .87]

Note. CI = confidence interval.

RTs were faster in the negative facts and negative gossip relative to the neutral condition, RTs in the positive conditions did not differ from the neutral condition, and while RTs were faster for negative compared to positive information, there was no main effect of trustworthiness or an interaction with valence (see Table 9 and 10).

Table 9

Means and confidence intervals for reaction times (in ms) for person judgments in experiment 2

Information	Neutral	Negative Facts	Negative Gossip	Positive Facts	Positive Gossip
M	879.25	825.21	818.73	866.67	885.33
95% CI	[854.88, 903.63]	[791.57, 858.85]	[797.24, 840.211]	[836.20, 897.14]	[849.86, 920.81]

Note. CI = confidence interval.

Table 10

Summary of statistical results for reaction times for person judgments in experiment 2

Source	*df*	*F*	*p*	η_p^2	η_G^2
Negative facts *vs*. Negative gossip *vs.* Neutral	2,46	7.56	.001	.25	.25
Negative Facts *vs*. Neutral	1,23	13.39	.004	.37	.37
Negative Gossip *vs.* Neutral	1,23	13.37	.004	.37	.37
Negative Facts *vs*. Negative Gossip	1,23	.11	1	.005	.005
Positive Facts *vs*. Positive Gossip *vs.* Neutral	2,46	.40	.67	.017	.017
Valence (Positive *vs.* Negative)	1,23	11.55	.002	.33	.16
Trustworthiness (Facts *vs.* Gossip)	1,23	.13	.73	.005	.002
Valence : Trustworthiness	1,23	1.19	.29	.049	.010

Event-Related Potentials

EPN. In the EPN time window between 200 and 300 ms no effects of negative or positive relative to neutral information were found. However, we also analyzed the time window from 300 to 350 ms since the EPN has been found slightly later for newly learned faces (up to 350 ms, *cf.* Luo et al., 2016; Suess et al., 2015; Xu et al., 2016; see Figure 1 in Supplemental Material).

Regarding the negative condition, no main effect of information was found between 200 and 300 ms, $F(2,46) = .047$, $p = .95$, $\eta_p^2 = .002$, $\eta_G^2 < .001$. Between 300 and 350 ms, the main effect did not reach significance, $F(2,46) = 2.87$, $p = .067$, $\eta_p^2 = .11$, $\eta_G^2 = .009$. In the positive condition, no main effect of information was found between 200 and 300 ms, $F(2,46) = 1.46$, p

= .24, η_p^2 = .060, η_G^2 = .004, but conditions differed between 300 and 350 ms $F(2,46) = 3.77$, p = .030, η_p^2 = .14, η_G^2 = .008. Analyses comparing positive facts to neutral information, $F(1,23)$ = 5.01, p = .11, η_p^2 = .18, η_G^2 = .007, and to positive gossip, $F(1,23) = 5.47$, p = .085, η_p^2 = .19, η_G^2 = .011, did not reach significance after Bonferroni corrections. Positive gossip did not differ from neutral information, $F(1,23) = .36$, $p = 1$, η_p^2 = .015, $\eta_G^2 < .001$.

LPP. Facts and gossip (negative and positive) elicited an enhanced positivity compared to neutral information (see Figure 5).

In the negative condition a main effect of information was found, $F(2,46) = 7.40$, p = .002, η_p^2 = .24, η_G^2 = .057. Separate analyses revealed an enhanced positivity for facts compared to neutral information, $F(1,23) = 14.73$, p = .003, η_p^2 = .39, η_G^2 = .074, as well as for gossip compared to neutral information, $F(1,23) = 9.67$, p = .015, η_p^2 = .30, η_G^2 = .061. There was no amplitude difference between negative facts and gossip, $F(1,23) = .050$, $p = 1$, η_p^2 = .002, $\eta_G^2 < .001$.

Concerning the positive condition a main effect of information was found, $F(2,46)$ = 5.69, p = .006, η_p^2 = .20, η_G^2 = .044. An enhanced positivity was found for positive facts compared to neutral information, $F(1,23) = 10.44$, p = .011, η_p^2 = .31, η_G^2 = .046, as well as for positive gossip compared to neutral information, $F(1,23) = 9.81$, p = .014, η_p^2 = .30, η_G^2 = .059. Positive facts and gossip did not differ, $F(1,23) = .072$, $p = 1$, η_p^2 = .003, $\eta_G^2 < .001$.

An analysis of valence (excluding the neutral condition) and trustworthiness, showed no main effect of valence, $F(1,23) = .16$, p = .69, η_p^2 = .007, $\eta_G^2 < .001$, no main effect of trustworthiness, $F(1,23) = .002$, p = .96, $\eta_p^2 < .001$, $\eta_G^2 < .001$, and no interaction of valence and trustworthiness, $F(1,23) = .13$, p = .72, η_p^2 = .006, $\eta_G^2 < .001$.

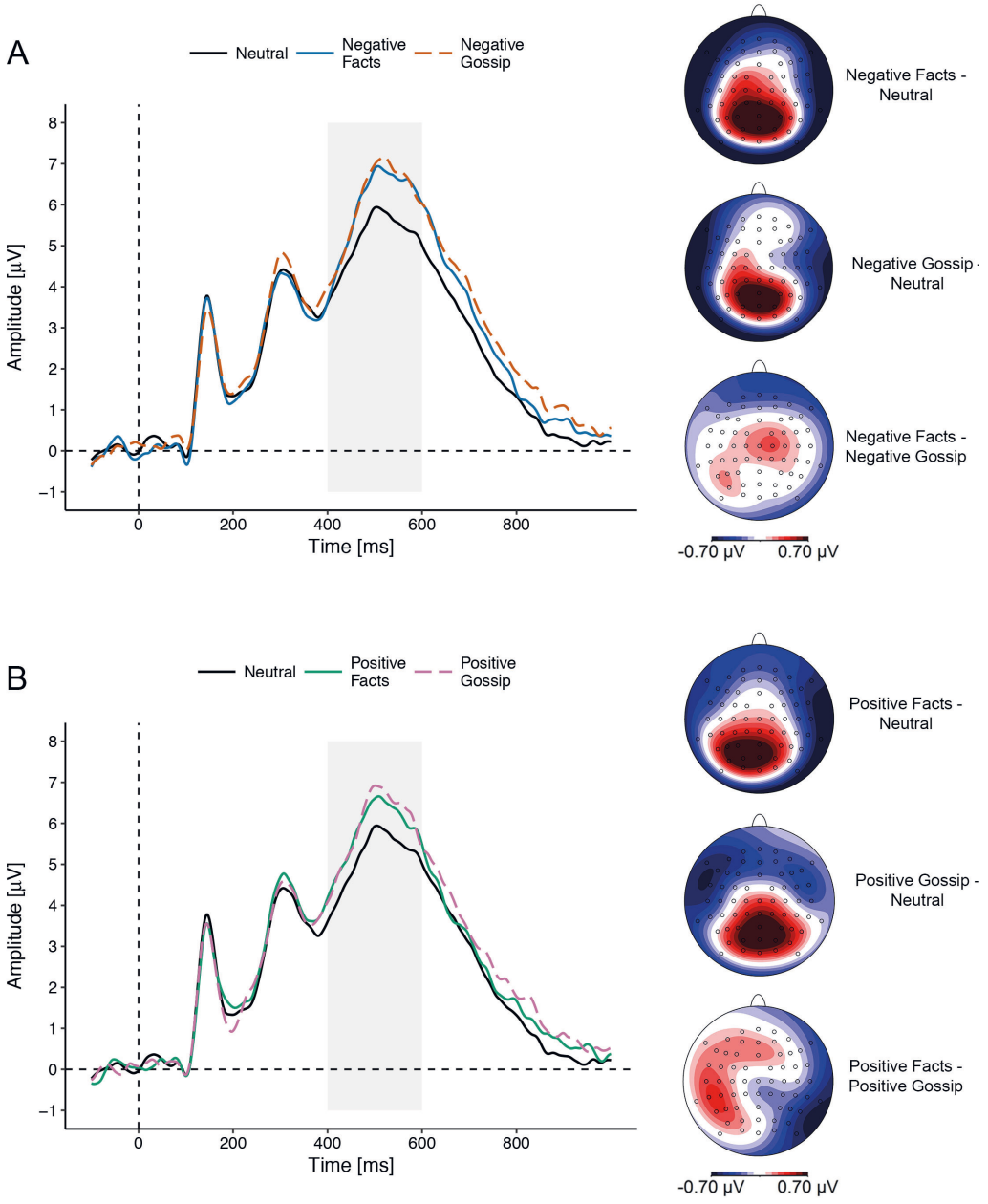

Figure 5. Grand average ERPs at the central-posterior site Pz show the LPP effects of person-related information in the person judgment task in experiment 2. Time point zero marks onset of face presentation. Scalp distributions show the effects as differences between conditions in the LPP time window between 400 and 600 ms (grey area). (A) Effects for negative information. (B) Effects for positive information.

Discussion

The aim of experiment 2 was to replicate the findings of experiment 1 with negative information and to test whether a similar pattern would be found for positive information or whether the trustworthiness of positive information would be taken into account, in contrast to negative information. This would be expected if participants ignore the trustworthiness to prioritize negative and potentially threatening information. Thus, for positive gossip we expected reduced effects of emotional evaluations reflected in explicit person judgments and LPP amplitudes. Emotion effects in the EPN were not expected to be modulated by trustworthiness.

Experiment 2 replicates the results of experiment 1 and in extension demonstrates that positive gossip also strongly affects spontaneous ratings of likability, explicit person judgments and LPP amplitudes reflecting evaluative processes. Again, none of these effects was modulated by the trustworthiness of the information, even though gossip was explicitly identified by the same participants as less trustworthy than facts. A statistically weak modulation of the EPN was restricted to a time window between 300 and 350 ms, and if anything, showed an unexpected tendency towards stronger effects for trustworthy information.

General Discussion

How people judge and emotionally evaluate others is influenced by what they know and what they hear about them. Here we show that person judgments are dominated by the social-emotional contents of person-related information, even when the information is clearly marked and understood as untrustworthy. Specifically, untrustworthy gossip strongly influenced spontaneous likability ratings and person judgments as well as brain responses in the LPP indexing emotional person evaluation.

Crucially, the gossip-like verbal information that was presented here clearly conveyed untrustworthiness and vagueness. Verbal qualifiers and constructions including e.g. “allegedly”, “supposedly”, “people assume”, and so forth put the contents of messages into perspective and weaken their meaning by indicating questionable reliability. In line with this, linguistic evidence shows that such expressions indeed change the meaning of verbal messages (e.g., Haertl, 2017; Schumacher, Brandt, & Weiland-Breckle, 2016). Furthermore, these expressions are commonly used in spoken and written form to indicate that information may not be truthful – and in professional journalistic contexts precisely such expressions are used to prevent prejudice and legal consequences of wrongly accusing and therefore defaming possibly innocent persons.

Besides the communicative and legal function of the verbal qualifiers employed here, we can demonstrate that the untrustworthiness was apprehended by our participants. The trustworthiness ratings revealed that participants differentiated between trustworthy and untrustworthy information and directly identified gossip as less trustworthy. We therefore conclude that the available verbal context information about the questionable trustworthiness was understood, but deliberately or involuntarily ignored during person evaluation. This finding dovetails with studies of the reliance on inaccurate information, even when it is easily identified and knowingly wrong (Hansen, Gerbasi, Todorov, Kruse, & Pronin, 2014; Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012; Rapp, 2016).

We had expected that person judgments based on untrustworthy compared to trustworthy information should affect us less, at least in the positive condition, if negative information is prioritized irrespectively as threatening. In contrast to these predictions, we found highly similar and robust effects when positive or negative social-emotional information was presented fact-like and when it was verbally marked – and recognized – as untrustworthy. The LPP findings in

combination with the behavioral results thus indicate that the knowledge about the untrustworthiness of person-related information does not automatically result in regulations of emotional responses in the service of arriving at accurate and fair judgments. Late aspects in the process of emotional appraisal and evaluation therefore seem to rely more on emotional contents for one's well-being, coping possibilities, and moral standards, even at the risk of possible misjudgments.

Limitations and Prospects for Future Research

Was our manipulation of trustworthiness too weak? The manipulation checks show that gossip was rated as significantly less trustworthy than facts. However, the mean difference seems comparatively small, raising the question whether the robust knowledge effects found in the gossip condition are due to a failure to induce a sufficiently strong manipulation. To address this point directly, we conducted additional Bayes factor hypothesis tests[2] (Wagenmakers, Marsman, et al., 2018) on the rating data of the manipulation checks and the person judgments, quantifying the relative evidence of the data in favor of the null hypothesis or the alternative hypothesis (see Supplemental Material page 12 for details). For the manipulation check in experiment 1, a Bayes factor of 42 indicates that it is 42 times more likely that facts were rated more trustworthy than was gossip, and for experiment 2 (across the positive and negative conditions) a Bayes factor of 47 indicates that it is 47 times more likely that facts were rated more trustworthy than was

[2] "The Bayes factor hypothesis test compares the predictive adequacy of two competing statistical models, thereby grading the evidence provided by the data on a continuous scale, and quantifying the change in belief that the data bring about for the two models under consideration" (Wagenmakers, Marsman, et al., 2018, p. 37)

gossip. Thus, we can consider the data of the manipulation checks as very strong evidence in favor of an effect of trustworthiness (for classification of Bayes Factors see Wagenmakers, Love, et al., 2018). We additionally estimated Bayes factors for the rating data in the person judgments that reveal that for both facts and gossip, judgments were over 100 times more likely to be negative (or positive in the case of positive information) than when based on neutral information (see Supplemental Material, Table S6). Concerning the direct comparison of facts and gossip we found inconclusive (experiment 1; Bayes factor of 1.3), and moderate evidence that facts and gossip did not indeed result in different judgments (experiment 2; Bayes Factor in favor of the null hypothesis of 4.6 for negative judgments, meaning it was 4.6 times more likely that there was no difference between negative facts and gossip, and of 3.7 for positive judgments, meaning it was 3.7 more likely that there was no difference between positive facts and gossip). Taken together, the Bayesian analyses reveal additional evidence that the trustworthiness effects are robust, suggesting that we have not simply failed to induce sufficiently strong effects.

Crucially, as discussed above, with our use of verbal markers to vary the trustworthiness of the information we chose a manipulation that reflects the actual use of such qualifiers in every day conversations (gossiping), and specifically in the news and social media. If our manipulation has no effect on person judgments and emotional responses to gossip, the frequent use and legal function of qualifiers as “allegedly” to prevent negative consequences of wrong accusations might be of questionable value.

One may also ask if the judgment task was engaging enough and if participants were lacking motivation to take the trustworthiness of the information under consideration. By including well-known filler faces and associated information, we created the overall impression that the persons and information encountered were existing, and thus that judgments were made

about real persons. It is a frequent real-life experience to read or hear information about unfamiliar people in many situations, and person judgments and evaluations are made deliberately and even happen spontaneously (e.g. Todorov, Gobbini, Evans, & Haxby, 2007). In such situations, without further motivation, we seem to care little about trustworthiness. The situation may be different when the motivation to care about the trustworthiness is enhanced (see discussion below).

It is also conceivable that our finding that person judgments and evaluations based on gossip are not tempered is related to a source monitoring deficiency (Johnson, Hashtroudi, & Lindsay, 1993). Specifically, participants may remember the (emotional) gist of the information, but not the trustworthiness-related qualifiers or alternatively, they may ignore the information already in the encoding phase. We cannot distinguish between these two alternatives based on our present data. However, future studies may investigate trustworthiness effects at encoding and recognition stages. Even if the precise mechanisms have yet to be described in full detail in future studies, our findings demonstrate that we strongly base our judgments on the emotional content while verbal qualifiers do not seem to have the often intended effects.

Last, in the judgment task, we used a categorical button-press answer format, which reduced artifacts during measuring ERPs, but does not enable nuanced judgments. One could argue that this forced participants into strong judgments. However, likability was measured on a scale with nuanced ratings and resulted in highly similar effects as for the judgments, suggesting that the answer format cannot explain the effects.

In summary, our results show that affective person judgments rely heavily on the emotional content of the information, while the reliability plays a minor role. The next step for future research is to think about what other factors could lead to a consideration of

trustworthiness. For example, motivation can be an important factor. Participants may be more inclined to consider trustworthiness when they know and care about the target of their judgments or more generally, when their judgments may have direct or indirect consequences. Another factor may be the arousal of the information and the social relevance. Some of the person information used in the present experiments were relatively extreme interpersonal behaviors (e.g. rape, saving someone's life; but also stealing, kindness; see Supplemental Material). It was important to use socially relevant information to be consistent with the effects found in past studies investigating person perception and evaluation (e.g. Abdel Rahman, 2011; Anderson et al., 2011; Bliss-Moreau et al., 2008; Wieser et al., 2014). Trustworthiness could have different effects depending on the intensity or sociality of the information, making this a topic for further investigation.

Conclusion

Our findings bear practical relevance. As in real life situations when confronted with social-emotional person-related information of varying levels of trustworthiness, participants were not instructed to actively suppress the emotional content or to contemplate the untrustworthiness of gossip but were free to use the available trustworthiness information to put their judgments into perspective. We demonstrate that person evaluation and person judgments – frequent activities in our daily social life and instances of everyday moral decisions (Helion & Ochsner, 2018; Hofmann et al., 2014), are strongly influenced by gossip, even when it is verbally marked, and can easily be identified as untrustworthy. Future research may target emotion regulation as possible strategy (Maroney & Gross, 2014) when emotional responses and biased judgments based on gossip cannot be prevented.

References

Abdel Rahman, R. (2011). Facing good and evil: Early brain signatures of affective biographical knowledge in face recognition. *Emotion*, *11*(6), 1397–1405. http://doi.org/10.1037/a0024717

Anderson, E., Siegel, E. H., Bliss-Moreau, E., & Barrett, L. F. (2011). The Visual Impact of Gossip. *Science*, *332*(6036), 1446–1448. http://doi.org/10.1126/science.1201574

Baum, J., Rabovsky, M., Rose, B. S., & Abdel Rahman, R. (2018). Repository: Clear judgments based on unclear evidence: Person evaluation is strongly influenced by untrustworthy gossip. Retrieved from osf.io/jqv2g

Berg, P., & Scherg, M. (1991). Dipole models of eye movements and blinks. *Electroencephalography and Clinical Neurophysiology*, *79*(1), 36–44.

Blechert, J., Sheppes, G., Di Tella, C., Williams, H., & Gross, J. J. (2012). See What You Think: Reappraisal Modulates Behavioral and Neural Responses to Social Stimuli. *Psychological Science*, *23*(4), 346–353. http://doi.org/10.1177/0956797612438559

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. http://doi.org/10.1037/1528-3542.8.4.479

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59. doi:10.1016/0005-7916(94)90063-9

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods*, *42*(1), 351–362. http://doi.org/10.3758/BRM.42.1.351

Ecker, U. K. H., Lewandowsky, S., Chang, E. P., & Pillai, R. (2014). The effects of subtle

misinformation in news headlines. *Journal of Experimental Psychology: Applied*, *20*(4), 323–335. http://doi.org/10.1037/xap0000028

Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Series in affective science.* (pp. 572–595). New York, NY, US: Oxford University Press.

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*(2), 175–191. http://doi.org/10.3758/BF03193146

Foti, D., & Hajcak, G. (2008). Deconstructing reappraisal: Descriptions preceding arousing pictures modulate the subsequent neural response. *Journal of Cognitive Neuroscience*, *20*(6), 977–988. http://doi.org/10.1162/jocn.2008.20066

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, *106*(1), 148–168. http://doi.org/10.1037/a0034726

Haertl, H. (in press). Name-informing and distancing ‘sogenannt’ (‘so-called’): Name mentioning and the lexicon-pragmatics interface. *Zeitschrift für Sprachwissenschaft, 37,* 139–169. Retrieved from https://ling.auf.net/ lingbuzz/003912

Hansen, K., Gerbasi, M., Todorov, A., Kruse, E., & Pronin, E. (2014). People Claim Objectivity After Knowingly Using Biased Strategies. *Personality and Social Psychology Bulletin*, *40*(6), 691–699. http://doi.org/10.1177/0146167214523476

Helion, C., & Ochsner, K. N. (2018). The Role of Emotion Regulation in Moral Judgment. *Neuroethics*, *11,* 297–308. http://doi.org/10.1007/s12152-016-9261-z

Herbert, C., Pauli, P., & Herbert, B. M. (2011). Self-reference modulates the processing of

emotional stimuli in the absence of explicit self-referential appraisal instructions. *Social Cognitive and Affective Neuroscience*, *6*(5), 653–661. http://doi.org/10.1093/scan/nsq082

Hertwig, R., & Engel, C. (2016). Homo Ignorans: Deliberately Choosing Not to Know. *Perspectives on Psychological Science*, *11*(3), 359–372. http://doi.org/10.1177/1745691616635594

Hinojosa, J. A., Méndez-Bértolo, C., & Pozo, M. A. (2010). Looking at emotional words is not the same as reading emotional words: Behavioral and neural correlates. *Psychophysiology*, *47*, 748–757. http://doi.org/10.1111/j.1469-8986.2010.00982.x

Hofmann, W., Wisneski, D. C., Brandt, M. J., & Skitka, L. J. (2014). Morality in everyday life. *Science*, *345*(6202), 1340–1343. http://doi.org/10.1126/science.1251560

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3–28. http://doi.org/10.1037//0033-2909.114.1.3

Kissler, J., Herbert, C., Winkler, I., & Junghöfer, M. (2009). Emotion and attention in visual word processing—An ERP study. *Biological Psychology*, *80*(1), 75–83. http://doi.org/10.1016/j.biopsycho.2008.03.004

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion*, *24*(8), 1377–1388. http://doi.org/10.1080/02699930903485076

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest*, *13*(3), 106–131. http://doi.org/10.1177/1529100612451018

Luo, Q. L., Wang, H. L., Dzhelyova, M., Huang, P., & Mo, L. (2016). Effect of Affective

Personality Information on Face Processing: Evidence from ERPs. *Frontiers in Psychology*, *7*, 1397. http://doi.org/10.3389/fpsyg.2016.00810

Maroney, T. A., & Gross, J. J. (2014). The Ideal of the Dispassionate Judge: An Emotion Regulation Perspective. *Emotion Review*, *6*(2), 142–151. http://doi.org/10.1177/1754073913491989

Mocaiber, I., Pereira, M. G., Erthal, F. S., Machado-Pinheiro, W., David, I. A., Cagy, M., et al. (2010). Fact or fiction? An event-related potential study of implicit emotion regulation. *Neuroscience Letters*, *476*(2), 84–88. http://doi.org/10.1016/j.neulet.2010.04.008

Molet, M., Kosinski, T., Craddock, P., Miguez, G., Mash, L. E., & Miller, R. R. (2016). Attenuating social affective learning effects with Memory Suppression manipulations. *Acta Psychologica*, *164*, 136–143. http://doi.org/10.1016/j.actpsy.2016.01.001

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review*, *108*(3), 483–522. http://doi.org/10.1037/0033-295X.108.3.483

Rabovsky, M., Stein, T., & Abdel Rahman, R. (2016). Access to Awareness for Faces during Continuous Flash Suppression Is Not Modulated by Affective Knowledge. *PLoS ONE*, *11*(4), e0150931. http://doi.org/10.1371/journal.pone.0150931

Rapp, D. N. (2016). The Consequences of Reading Inaccurate Information. *Current Directions in Psychological Science*, *25*(4), 281–285. http://doi.org/10.1177/0963721416649347

Rellecke, J., Sommer, W., & Schacht, A. (2012). Does processing of emotional facial expressions depend on intention? Time-resolved evidence from event-related brain potentials. *Biological Psychology*, *90*(1), 23–32. http://doi.org/10.1016/j.biopsycho.2012.02.002

Sabatinelli, D., Keil, A., Frank, D. W., & Lang, P. J. (2013). Emotional perception: Correspondence of early and late event-related potentials with cortical and subcortical functional MRI. *Biological Psychology*, *92*(3), 513–519. http://doi.org/10.1016/j.biopsycho.2012.04.005

Schacht, A., & Sommer, W. (2009). Emotions in word and face processing: Early and late cortical responses. *Brain and Cognition*, *69*(3), 538–550. http://doi.org/10.1016/j.bandc.2008.11.005

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Series in affective science.* (pp. 92–120). New York, NY, US: Oxford University Press.

Schumacher, P. B., Brandt, P., & Weiland-Breckle, H. (2016). Online processing of "real" and “fake”: The cost of being too strong. In E. Castroviejo, L. McNally & G. W. Sassoon (Eds.), *The Semantics of Gradability, Vagueness, and Scale Structure: Experimental Perspectives*. (pp. 1-20). Heidelberg: Springer Verlag.

Schupp, H. T., Junghöfer, M., Weike, A. I., & Hamm, A. O. (2003). Emotional Facilitation of Sensory Processing in the Visual Cortex. *Psychological Science*, *14*(1), 7–13. http://doi.org/10.1111/1467-9280.01411

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*. http://doi.org/10.1037/emo0000305

Suess, F., Rabovsky, M., & Abdel Rahman, R. (2015). Perceiving emotions in neutral faces: expression processing is biased by affective person knowledge. *Social Cognitive and Affective Neuroscience*, *10*(4), 531–536. http://doi.org/10.1093/scan/nsu088

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

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., … Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, *25*(1), 58–76. http://doi.org/10.3758/s13423-017-1323-7

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., … Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*(1), 35–57. http://doi.org/10.3758/s13423-017-1343-3

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. http://doi.org/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. *Frontiers in Psychology*, *07*(1368), 1446. http://doi.org/10.1016/j.biopsycho.2014.11.012

Yoder, K. J., & Decety, J. (2014). Spatiotemporal neural dynamics of moral judgment: A high-density ERP study. *Neuropsychologia*, *60*, 39–45. http://doi.org/10.1016/j.neuropsychologia.2014.05.022

Acknowledgments

We thank Luisa Balzus, Marie Borowikow, Cornelius Braun, Anastasia Matsneva, Johannes Rost, and Jule Schmidt for their help during data collection and Guido Kiecker for technical assistance.

OXFORD

Social Cognitive and Affective Neuroscience, 2021, 280–291

doi:10.1093/scan/nsaa164
Advance Access Publication Date: 4 December 2020
Original Manuscript

Emotional news affects social judgments independent of perceived media credibility

Julia Baum[1,2] and Rasha Abdel Rahman[1,2]

[1]Faculty of Philosophy, Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Berlin 10099, Germany, and [2]Faculty of Life Sciences, Department of Psychology, Humboldt-Universität zu Berlin, Berlin 10099, Germany

Correspondence should be addressed to Julia Baum, Institut für Psychologie, Humboldt-Universität zu Berlin, Unter den Linden 6, Berlin 10099, Germany. E-mail: julia.baum@hu-berlin.de.

Abstract

How does the credibility we attribute to media sources influence our opinions and judgments derived from news? Participants read headlines about the social behavior of depicted unfamiliar persons from websites of trusted or distrusted well-known German news media. As a consequence, persons paired with negative or positive headlines were judged more negative or positive than persons associated with neutral information independent of source credibility. Likewise, electrophysiological signatures of slow and controlled evaluative brain activity revealed a dominant influence of emotional headline contents regardless of credibility. Modulations of earlier brain responses associated with arousal and reflexive emotional processing show an effect of negative news and suggest that distrusted sources may even enhance the impact of negative headlines. These findings demonstrate that though we may have distinct perceptions about the credibility of media sources, information processing and social judgments rely on the emotional content of headlines, even when they stem from sources we distrust.

Key words: news media trust; misinformation; evaluative learning; emotional person knowledge; social judgments; event-related potentials

In times of massive online communication, news and information from various sources spread rapidly, shaping personal opinions as well as public debates (Vosoughi *et al.*, 2018). Aside from well-vetted news, intentionally or unintentionally spread misinformation, 'fake news' and 'alternative facts' have gained influence (Lazer *et al.*, 2018; Lewandowsky *et al.*, 2012). Despite the potentially detrimental effects of misinformation and their increasing prevalence in (social) media and political discourse, research on the consequences of being exposed to misinformation is scant, and little is known about the behavioral and neural correlates of processing information of questionable veracity (Baum *et al.*, 2018). Experimental evidence revealing insights into the cognitive mechanisms can be vital to a comprehensive understanding of how we are affected by information from media (as argued, e.g. by Lazer *et al.*, 2018; Vosoughi *et al.*, 2018 and Aral and Eckles, 2019).

One resource-efficient and fast heuristic to assess the veracity of news is to consider the credibility of the source. Indeed, recent evidence suggests that we trust or distrust media sources based on criteria as familiarity, likability, social endorsement and reputation, and laypeople's credibility assessments align with those of professional fact checkers (Metzger and Flanagin, 2013; Pennycook and Rand, 2018, 2019). However, despite our ability to evaluate the credibility of a source, little is known about the impact of such assessments on the cognitive processes underlying social judgments and decisions. The aim

Received: 25 July 2020; **Revised:** 1 November 2020; **Accepted:** 3 December 2020

of the current study is to investigate the later (and possibly memory-related) consequences of having been exposed to news from various sources. Specifically, we asked how the perceived credibility of existing and well-known news sources affects subsequent information processing and social judgments based on person-related negative or positive headlines. We extracted event-related brain potentials (ERPs) from the electroencephalogram (EEG) to localize the effects and interactions of social-emotional information and source credibility at early reflexive and later more controlled processing stages to gain insight into the underlying cognitive mechanisms and brain signatures (Figure 1 for the study phases).

When we are exposed to news, we are confronted with verbal information (Figure 1, Phase 1). Emotional person-related verbal

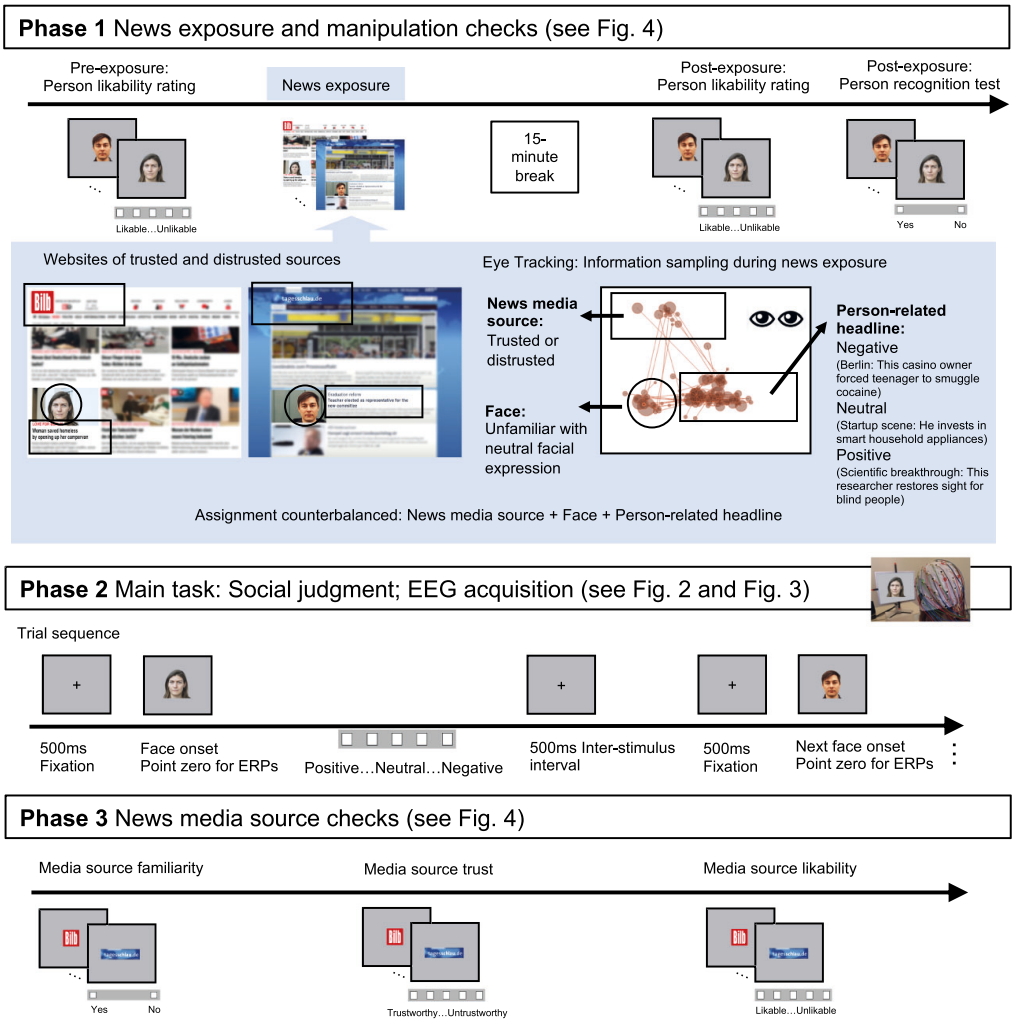

Fig. 1. Overview of the well-controlled experimental study design with three phases. In Phase 1, participants were exposed to experimental but authentic websites of existing and widely distributed mainstream German news media (e.g. 'Tagesschau' and 'Bild') that were selected based on their pre-rated high or poor credibility. English-speaking analogies may be e.g. 'BBC', 'Fox News', 'The Guardian', 'The New York Times', 'Daily Mirror' or 'The Sun'. Each website presented the news media source logo, the face and the headline containing negative, positive or neutral emotional person-related information; all other details were blurred (in the experiment original layouts, logos and fonts were used). To enhance authenticity, we added news reports about well-known persons as fillers. The assignment of unfamiliar faces to conditions was counterbalanced: while one participant was exposed to each face only in one context condition, the faces were presented equally often in each condition across participants. An additional eye-tracking experiment with different participants verified the sampling of source information during news exposure (shown here: example data of one participant for one website, lines represent saccades, points represent fixations and point magnitude represents their duration). To check whether the news exposure manipulation was successful, we subsequently tested whether the faces were reliably recognized and how likable participants found each person before and after news exposure. In Phase 2, the main experimental task followed, in which the faces were presented in isolation and the EEG was registered, while participants judged the depicted persons based on the information they had been exposed to (social judgment). Just as it is typically the case when reading news headlines, participants were not explicitly instructed to consider the credibility of the source. Instead, they were asked to make their judgment based on the information from Phase 1. In Phase 3 (after the main task), participants rated the familiarity, likability and credibility of the news media sources as an additional manipulation check.

information—even when minimal like in headlines—can change the affective value of people by mechanisms of verbal evaluative learning (also referred to as evaluative conditioning) as well as by attributional or propositional processes that may additionally take into account the relevance or truth-value of the information in its context (Bliss-Moreau *et al.*, 2008; Mattarozzi *et al.*, 2014; Ferrari *et al.*, 2020; for a general review, see De Houwer, Van Dessel, Moran, 2020). Some evidence of potential neural underpinnings of person-related verbal evaluative learning suggest that while emotional information may not affect very early visual processing (but see Galli *et al.*, 2006), it can affect early and later conceptual processing that may rely on both implicit and explicit memory of the information (Kissler and Strehlow, 2017; Junghöfer *et al.*, 2016 and see introduction of ERP effects below). Yet, research on how these effects are modulated by the veracity of the information is scarce (Baum *et al.*, 2018).

What are the expected consequences of having been exposed to emotional news from trusted and distrusted sources on social judgments (Figure 1, Phase 2)? The family of dual-process theories distinguishes between two separate systems or interactive processes related to fast, impulsive, spontaneous and automatic processing on the one hand and slower, intentional and controlled processing on the other (e.g. Kahneman, 2003; Strack and Deutsch, 2004; Gawronski and Bodenhausen, 2006; Cunningham and Zelazo, 2007 and Lieberman, 2007). This concept also relates to models of recognition and memory distinguishing faster and slower retrieval, with slower processes retrieving additional context and source information that may be stored unitized or separately (for a review, see Yonelinas, 2002). For the memory-related processing in Phase 2, this suggests that our cognitive system initially spontaneously processes the emotional content of the headlines associated with the person irrespective of the credibility of the source, whereas later, more controlled processes should result in evaluations that take the credibility of the source into account, resulting in social judgments that are qualified according to the presumed credibility.

With respect to emotion processing, appraisal theories (Scherer, 2001; Ellsworth and Scherer, 2003) assume that stimuli are initially checked for a coarse detection of emotional salience, intrinsic pleasantness and arousal. This is followed by assessments regarding implications for the observer's well-being, coping possibilities and evaluations of the normative significance, like the compatibility with moral standards. This may also include the truth-value of information. Concerning the impact of news, and in analogy to dual-process theories, emotional contents and source credibility should be processed at different points in time. While early emotional responses should be influenced only by the emotional content of headlines, later more controlled processes should take source credibility into account.

In ERPs, fast and early processing has been related to an enhanced early posterior negativity (EPN) at about 200–300 ms at occipito-temporal brain regions that indexes reflexive and arousal-related emotional processes (e.g. Junghöfer *et al.*, 2001; Schupp *et al.*, 2003, 2004; Kissler *et al.*, 2007). At later stages, an enhanced late positive potential (LPP) at about 400–600 ms at centro-parietal regions is associated with elaborate and reflective processing (Schupp *et al.*, 2004; Schacht and Sommer, 2009a; Sabatinelli *et al.*, 2013). Both components are sensitive to verbal affective person-related information associated with faces via verbal evaluative learning (for instance, EPN: Abdel Rahman, 2011; Kissler and Strehlow, 2017; Wieser *et al.*, 2014; Suess *et al.*, 2015; Junghöfer *et al.*, 2016; Luo *et al.*, 2016; Xu *et al.*, 2016; LPP: Luo *et al.*, 2016; Baum *et al.*, 2018). Crucially, the LPP is sensitive to additional information such as context and relevance, putting emotional contents into perspective (Schacht and Sommer, 2009b; Herbert *et al.*, 2011, 2013; Blechert *et al.*, 2012; Rellecke *et al.*, 2012; Schindler *et al.*, 2019), whereas the EPN is relatively independent of task demands and the relevance of emotional contents in a given context (Schacht and Sommer, 2009b; Herbert *et al.*, 2011, 2013). It is noteworthy that this evidence of additional contextual influences on ERPs comes from studies testing the effects of emotional information immediately, while there is scarce evidence of such contextual effects on later consequences (Baum *et al.*, 2018). We expected that the EPN is mainly sensitive to the emotional content of the headlines irrespective of source credibility, whereas emotion effects in LPP amplitudes should be modulated by source credibility, with reduced amplitudes for distrusted sources.

To summarize, based on dual-process theories distinguishing fast impulsive and slower more controlled processes, we expected that early processing of faces associated with emotional *vs* neutral headlines from trusted and distrusted sources should be modulated only by effects of emotion, whereas later controlled evaluation should take source credibility into account, resulting in tempered social judgments. This modulation may be primarily found for positive headlines if negative information is prioritized as protection against potential threat (cf. Baum *et al.*, 2018). The present study was preregistered under the OSF (Open Science Framework) (osf.io/scbgq).

Method

Participants

The sample size was preregistered and planned according to the requirements of the counterbalancing and based on power analyses, see Supplementary Data page 1. The final data set consisted of 30 participants [$M_{age} = 25$ (s.d. = 5.36), 25 females, all right-handed]. Four participants were excluded (one was familiar with face databases, two rated the trustworthiness equal across sources and one did not acquire person-related information) and replaced with new participants. Participants were compensated in the 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.

Materials

Websites of news media combined source, face and headline (for example Figure 1, Phase 1). We edited each colored face photograph onto a natural background (e.g. street scene and wall), inserted it onto the website and changed the headline via source code, keeping the characteristic font (with font size kept similar across media sources). Thus, we maintained the distinctive layout of the media sources while experimentally manipulating the content, since the layout and visual design of websites play an important role in assessing the credibility of a source (Metzger *et al.*, 2013). In Phase 1, website screenshots were displayed full screen and showed the prominent logo on top of the page, the face and the headline, while all other details were blurred. For the news exposure, 24 unfamiliar faces were equally assigned to neutral, negative and positive headlines, with counterbalanced assignment across participants. The assignment of faces and headlines to media sources was also counterbalanced across participants, with 12 target faces appearing in trusted sources

and 12 faces in distrusted sources, resulting in four target faces in each condition of the 3 × 2 design. Affective information for eight well-known filler faces referred to recent news about them, and the assignment of headlines was fixed for all participants.

News media sources were selected based on pre-ratings of credibility and familiarity with a different group of German participants [$N = 38$, 33 females, $M_{age} = 26$ (s.d. = 4.69), all students]. The pre-rating tested 35 German news media sources, including well-known, less-well-known and highly partisan sources. The rating scale was from 3 (very credible) to −3 (not credible). We selected the four sources rated as most credible ($M = 1.77$, 95%CI [1.57, 1.97]), and the four rated as least credible [$M = -1.64$, 95%-CI (−1.92, −1.37)], all highly familiar (familiar = 1, unfamiliar = 0; $M = 0.98$ for trusted and distrusted). The selected sources were 'Tagesschau', 'Sueddeutsche', 'Zeit Online', 'Frankfurter Allgemeine', 'Bild', 'B.Z. Berlin', 'Bunte' and 'Gala'. Credibility ratings were significantly higher for trusted than for distrusted sources, $t(37) = 14.83$, $P < 0.001$. Colored screenshots of the sources' logos were presented in similar size in the media source ratings of the current experiment (2.7 × 3.5 cm).

Face stimuli were colored frontal portraits of 24 unfamiliar faces with neutral facial expressions, presented on a gray background during the main task and manipulation checks (2.7 × 3.5 cm, viewing distance 70 cm; from multiple databases, see Supplementary Data page 14). Eight familiar filler faces (e.g. Emma Watson and Harvey Weinstein) were added to make the target persons' existence credible.

Headlines describing social behavior were either neutral, negative or positive (for all headlines see Supplementary Table S20). Pre-ratings with different participants confirmed their valence and showed that positive and negative headlines were equally more arousing than neutral headlines (see Supplementary Data, page 14).

Procedure

The procedure entails three phases (Figure 1) as a variant of a well-established design (cf. Abdel Rahman, 2011; Suess *et al.*, 2015; Baum *et al.*, 2018). In Phase 1, the experiment started with a person likability rating of all faces on a 5-point scale (pre-exposure rating). Response buttons were placed in front of participants. Then the news exposure followed. Participants were instructed as follows (here translated from German): 'You now receive information of various kinds about these people, taken from media reports. Unrelated content and details remain unrecognizable. Please read the information carefully'. Each trial started showing the website—which was blurred except for the logo of the media source—for one second. For the remaining 5 s, the logo, the face and the headline were unblurred. Websites were presented in blocks of eight, including all experimental conditions and two fillers. Each website was presented five times in total (160 trials in total). To keep participants engaged with the task, they occasionally answered short yes-or-no questions about the persons, e.g. 'Is the behavior of this person common?' (asked in ~22% of the trials of Phase 1). After completion of the news exposure, participants had a 15-minute break. Phase 1 ended with a post-exposure likability rating (see earlier) and a recognition test as manipulation checks. In the recognition test, participants decided whether a face had been encountered in the news exposure or not (this included 32 additional unfamiliar filler faces).

In Phase 2, the EEG was recorded while a social judgment task was employed as the main task. Participants judged how negative, neutral or positive the depicted person was based on information acquired in Phase 1. Participants judged on a 5-point scale, enabling them to nuance their answers between neutral and negative/positive. To enhance the signal-to-noise ratio necessary for the EEG data quality, the task was repeated 20 times block-wise, separated by breaks, resulting in 80 trials per condition (excluding fillers). Participants were told that the repetition of the task is a technical necessity for EEG measurements. Trials started with a 500 ms pre-stimulus fixation cross and had a 500 ms inter-trial interval. Faces were presented until response or for a maximum of 3 s.

Phase 3 entailed manipulation checks of the media sources. First, participants saw the logos and were asked if they knew the sources. Then they rated how trustworthy they consider each source, on a 5-point scale from trustworthy to untrustworthy while the EEG was recorded. The trust rating was repeated 10 times, resulting in 40 trials per condition and logos were presented until response. At last, participants were asked to rate how likeable they find each media source. This rating was included because likability may not necessarily be equivalent to credibility (e.g. one may enjoy reading gossip papers, without trusting its contents).

The direction of scales was counterbalanced, i.e. there were two versions, in version one five buttons ranged from positive (left) to negative (right), and in version two the buttons ranged from negative (left) to positive (right). This was consistent for all tasks and phases, i.e. very likeable, positive, yes and very credible on the left for version one and vice versa for version two. After the experiment, participants were asked to reproduce the contents of the headline about each person to check if they remembered the broad information. Phase 1 lasted 30 minutes and Phases 2 and 3 together lasted 40 minutes, and participants were compensated for all time spent at the lab.

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 5 kΩ. EEG data were 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, which 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 passband edge [zero-phase FIR (Finite Impulse Response) 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 (Brain Electrical Source Analysis) (Berg and 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 ±200 μV, changing >50 μV between samples or >200 μV within single epochs or containing baseline drifts. Error- and artifact-free EEG data were 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 79 trials per condition remained (range: 73–80) and in each condition

98% of trials were kept overall (neutral-trusted 2357, neutral-distrusted 2364, negative-trusted 2350, negative-distrusted 2355, positive-trusted 2363 and positive-distrusted 2362). Trials where no judgment was given were excluded (in the social judgment task there were 33 out of 14 400).

Data analysis

ERP analyses focus on two regions of interest (ROIs), the EPN (at electrode sites PO7, PO8, PO9, PO10, TP9, TP10, 200–350 ms after face stimulus onset) and the LPP component (Pz, CPz, POz, P3, P4, 400–600 ms), based on previous findings of emotional stimulus content (e.g. Schupp *et al.*, 2003) and affective information (e.g. Abdel Rahman, 2011 and Baum *et al.*, 2018). To explore effects occurring during early visual face processing, we additionally analyzed the P100 (PO3, PO4, O1, O2, 100–150 ms), and the N170 (P7, P8, PO7, PO8, 150–200 ms), based on previous findings (e.g. Abdel Rahman and Sommer, 2012). P100 and N170 results are available in the Supplementary Tables S7–S9. Amplitudes were averaged over ROIs and time windows on single-trial level.

We used mixed-effects regression models on single-trial data of behavioral measures and ERPs (Frömer *et al.*, 2018). For continuous dependent variables we used linear mixed models (LMMs; Bates *et al.*, 2015b: 'lme4' v.1.1-17 in R) and tested the significance of fixed effects coefficients (P-value < 0.05) by Satterthwaite approximation ('summary' function of 'lmerTest' v.3.0-1, Kuznetsova *et al.*, 2017). For ordinal dependent variables we used cumulative link mixed models fitted with Laplace approximation (CLMMs; ordinal v.2019.12-10, Christensen, 2019). For each dependent variable, the model was specified with fixed effects for the experimental factors 'headline content' (negative, positive, 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. Thus 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, Benjamini & Hochberg, 1995; Tables 2 and 4). We report point estimates (*b*), 95% confidence intervals (CIs) for LMMs, standard errors, *t*-values for LMMs, *z*-values for CLMMs and *P*-values for the fixed effects coefficients. Data and code can be accessed online (osf.io/dwesx/).

Results

Effects of emotional news on information processing and social judgments (Phase 2)

Behavioral results. Persons associated with negative headlines were judged as more negative relative to persons associated with neutral headlines, and judgments based on negative headlines were faster than when based on neutral headlines (Table 1 and Figure 2). Source credibility did not modulate the negative headline effects in judgment decisions and latencies (Table 1). Unexpectedly, social judgments based on negative

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 judgment decisions and latencies were estimated in separate mixed models, and fixed effects were coded as repeated contrasts according to our hypotheses

	Social judgment decisions				Social judgment latencies [−1000/latency(ms)]			
Coefficient	*b*	SE	z	*P*	*b* (95% CI)	SE	*t*	*P*
Intercept (grand mean)					−1.31 (−1.37 to −1.25)	0.03	−44.45	<0.001
Source credibility (trusted *vs* distrusted)	−0.01	0.39	−0.02	0.984	0.01 (−0.01 to 0.04)	0.01	0.84	0.410
Negative headline content (Neg *vs* Neu)	−8.20	0.75	−11.00	<0.001	−0.13 (−0.18 to −0.07)	0.03	−4.69	<0.001
Source credibility × negative headline content	−0.28	1.11	−0.26	0.799	−0.02 (−0.08 to 0.04)	0.03	−0.65	0.521
Positive headline content (Pos *vs* Neu)	4.93	0.61	8.07	<0.001	−0.06 (−0.10 to −0.02)	0.02	−3.04	0.004
Source credibility × positive headline content	−0.05	0.46	−0.12	0.908	−0.02 (−0.08 to 0.05)	0.03	−0.55	0.583
Model formula	Decision ~ headline content × source credibility + (S + Neg + S × Neg + Pos + S × Pos || subject) + (S + Neg + S × Neg + Pos + S × Pos || face)				Latency ~ headline content × source credibility + (S + Neg + S × Neg + Pos + S × Pos || subject) + (S + Neg + S × Neg + Pos + S × Pos || face)			

Double bars in random effects terms represent correlation parameters set 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.
CLMM threshold coefficients for social judgment decisions: −2|−1: $b = -3.9$, SE = 0.07, $z = -57.9$; −1|0: $b = -1.75$, SE = 0.05, $z = -33.2$; 0 × 1 $b = 2.0$, SE = 0.05, $z = 40.4$; 1|2: $b = 5.1$, SE = 0.07, $z = 74.3$. LMMs yield the same pattern of results for ordinal dependent variables as CLMMs, see Supplementary Table S2 for LMM results of ordinal dependent variables treated as continuous variables.

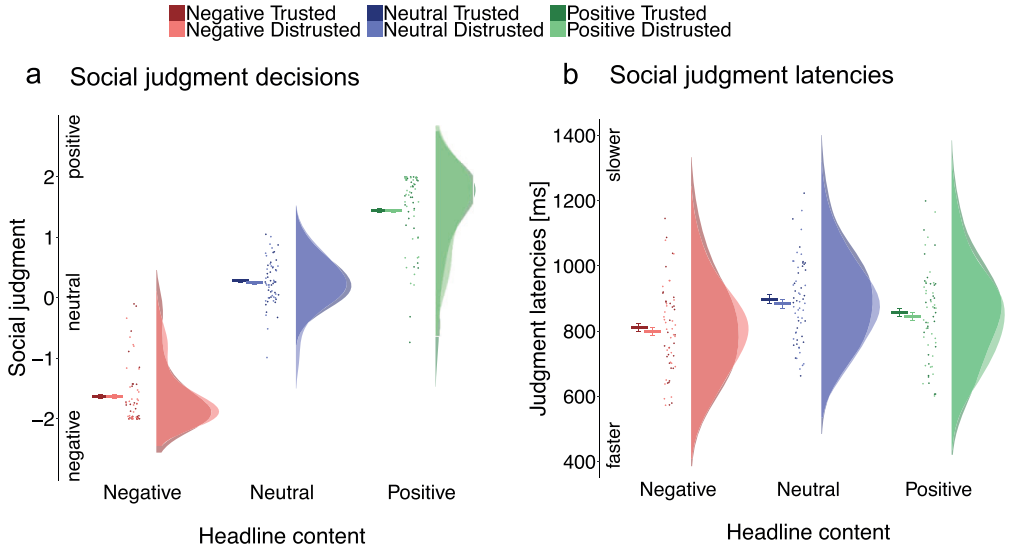

Fig. 2. In Phase 2, social judgment was performed as the main task to investigate the effects of emotional news and source credibility. Behavioral results show that (a) persons were judged based on emotional headline content, whereas source credibility had no influence. (b) Judgments based on emotional headlines were faster than neutral, but not tempered by source credibility. Raincloud plots (Allen *et al.*, 2019) show means and 95% CIs calculated with the 'summarySEwithin' function (Morey, 2008) on single trial data, with points and distributions for data aggregated by subject.

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

	Social judgment decisions				Social judgment latencies [−1000/latency(ms)]			
Contrast	*b*	SE	*z*	*P*	*b*	SE	*t*	*P*
Trusted: Neg *vs* Neu	−8.35	0.93	−8.98	<0.001	−0.14	0.03	−4.39	<0.001
Distrusted: Neg *vs* Neu	−8.06	0.93	−8.67	<0.001	−0.12	0.03	−3.74	<0.001
Trusted: Pos *vs* Neu	4.91	0.65	7.50	<0.001	−0.07	0.03	−2.71	0.011
Distrusted: Pos *vs* Neu	4.96	0.65	7.60	<0.001	−0.05	0.03	−2.00	0.049

vs neutral headlines were more negative and faster for both, trusted and distrusted sources (Table 2).

For positive headlines, social judgments were more positive and also faster compared to neutral headlines (Table 1). These effects were not modulated by source credibility (Table 1). Social judgments of positive *vs* neutral headlines were more positive and faster for trusted and distrusted sources (Table 2).

Post hoc (non-preregistered), we included repetition as a covariate to test whether social judgments and their latencies were biased toward focusing on emotional contents by repeating the task, which was necessary to ensure EEG data quality. The three-way interactions were not significant (all *t*-values < |.9|, all *P*-values > 0.4; see Supplementary Table S3). Moreover, testing only the first judgments per face (task was repeated block wise) resulted in the same pattern (Supplementary Table S4). We conclude that repetition did not change the result pattern.

Event-related brain potentials

EPN. To investigate relatively fast and reflexive emotional processing we focused on the EPN component. Negative compared to neutral headlines elicited an enhanced negativity, and there was a trend for an interaction with source credibility (please see Table 3). The EPN effect of negative headlines was enhanced for distrusted sources, but absent for trusted sources (Table 4 and Figure 3A,C).

For positive headlines, we found no EPN effect for positive compared to neutral headlines, no interaction with source credibility (Table 3) and no EPN effects nested in trusted or distrusted sources (Table 4 and Figure 3B,C).

LPP. To investigate more controlled evaluative processing, we tested effects in the later LPP component. For negative headlines, we found an enhanced LPP compared to neutral headlines and no interaction with source credibility (Table 3). Negative information from both, trusted and distrusted sources elicited LPP effects (Table 4 and Figure 3A,C).

For positive headlines, the LPP was enhanced compared to neutral headlines and there was no interaction with source credibility (Table 3). Positive information from trusted and distrusted sources elicited LPP effects (Table 4 and Figure 3B,C).

Post hoc (non-preregistered), we included judgment latencies as a covariate to account for motor responses in the LPP results. This did not change the effects of predictors, and three-way interactions were not significant (all *ts* < 1, all *ps* > 0.3; see Supplementary Table S6). We cannot fully exclude the possibility of motor-response or -preparation influences. Yet, we

Table 3. LMM summary statistics show effects of source credibility, negative and positive headline content and their interactions on ERPs 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 LMMs and fixed effects were coded as repeated contrasts according to our hypotheses

	EPN				LPP			
Coefficient	*b* (95% CI)	SE	*t*	*P*	*b* (95% CI)	SE	*t*	*P*
Intercept (grand mean)	2.41 (1.33 to 3.49)	0.55	4.36	<0.001	4.64 (3.88 to 5.41)	0.39	11.85	<0.001
Source credibility (trusted *vs* distrusted)	−0.02 (−0.23 to 0.18)	0.11	−0.23	0.819	0.10 (−0.11 to 0.31)	0.11	0.95	0.353
Negative headline content (Neg *vs* Neu)	−0.29 (−0.50 to −0.08)	0.11	−2.65	0.014	1.13 (0.80 to 1.45)	0.17	6.79	<0.001
Source credibility × negative headline content	0.42 (0.01 to 0.84)	0.21	2.00	0.056	0.36 (−0.06 to 0.78)	0.21	1.69	0.101
Positive headline content (Pos vs. Neu)	−0.11 (−0.32 to 0.09)	0.10	−1.09	0.287	0.50 (0.23 to 0.77)	0.14	3.60	0.001
Source credibility × positive headline content	0.14 (−0.33 to 0.61)	0.24	0.59	0.559	0.21 (−0.29 to 0.71)	0.26	0.83	0.414
Model formula	EPN ∼ headline content × source credibility + (S + Neg + S × Neg + Pos + S × Pos || subject) + (S + Neg + S × Neg + Pos + S × Pos || face)				LPP ∼ headline content × source credibility + (S + Neg + S × Neg + Pos + S × Pos || subject) + (S + Neg + Pos + S × Pos || face)			

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.

consider motor response confounds unlikely because first, all trials involved motor responses (Luck, 2014) and second, latency differences were taken into account in the model. Thus, mostly unsystematic or nonlinear motor-response-related differences could have affected the LPP.

News exposure and manipulation checks (Phase 1)

We manipulated headline content and news media credibility during news exposure and demonstrate that these manipulations were successful (Figure 4A,B). Pre-exposure person likability ratings were on average neutral (Supplementary Tables S10–S12), whereas after exposure persons were disliked when associated with negative headlines and liked when associated with positive headlines [$b = -1.52$, 95% CI (−1.73, −1.31), $t = -13.96$, $P < 0.001$ and $b = 0.78$, 95% CI (0.61, 0.95), $t = 9.01$, $P < 0.001$, respectively]. Source credibility did not modulate likability ratings ($ts < |.97|$, $ps > 0.3$). In the post-exposure recognition test, faces were successfully recognized across conditions, M = 97.3%. There were no effects of headline or source on accuracy (Supplementary Tables S13 and S14).

We conducted an additional eye-tracking experiment with different participants [N = 12, $M_{age} = 25$ (s.d. = 7.93), 8 females] to check whether participants acknowledge the media source during news exposure, without having been explicitly instructed (see 'Procedure' section). One-sample *t*-tests confirmed that per face presented in the website context, the source fixation durations and frequencies on the source logo were above zero [$M = 896$ ms, 95% CI (440,-); $t(11) = 3.53$, $P = 0.002$, $d = 1.02$ and $M = 4.1$, 95% CI (2.2,-); $t(11) = 3.93$, $P = 0.001$, $d = 1.14$; Figure 1 and Supplementary Data page 11]. Furthermore, we tested if the blurred layout in itself provides cues of the media source. In a separate task after news exposure, participants assigned screenshots of websites where the logo had been removed for one of two sources (correct media source *vs* logo of a different source from the other credibility condition). Ninety percentage of the layouts were correctly identified [$M = 0.90$, 95% CI (0.86,-), $t(11) = 40.58$, $P < 0.001$, $d = 11.71$].

News media source checks (Phase 3)

All participants were familiar with all media sources. Distrusted sources were rated as untrustworthy and less likable, whereas trusted sources were rated as trustworthy and likable [source credibility effect in trust ratings: $b = 3.02$, 95% CI (2.66, 3.38), $t = 16.64$, $P < 0.001$ and in likability ratings: $b = 2.56$, 95% CI (2.09, 3.02), $t = 10.80$, $P < 0.001$; see Figure 4C,D and Supplementary Data page 11f].

Discussion

Here we show that emotional person-related news headlines strongly affect subsequent information processing and social judgments irrespective of whether the source is perceived as credible or not. Emotional contents of headlines determined social judgments and affected slow evaluative brain responses in the LPP component known to be sensitive to context information and deliberate control. Crucially, none of these effects was modulated by source credibility, suggesting that headlines in news media may have an even stronger than expected influence on information processing and social judgments. Indeed, even if we assume that there are subtle traces of source credibility modulations that are difficult to detect, the fact remains that headlines from distrusted sources induce strong and robust effects of emotional information on social judgments.

Fast emotional brain modulations in the EPN component associated with arousal and sensation-related reflexive processing were modulated by emotional headline content and show furthermore that, if anything, distrusted sources may even enhance, instead of reduce, the impact of negative compared to neutral headlines. Please note however that this early interaction of headline content and source credibility was not predicted and the interaction was only marginally significant, even though

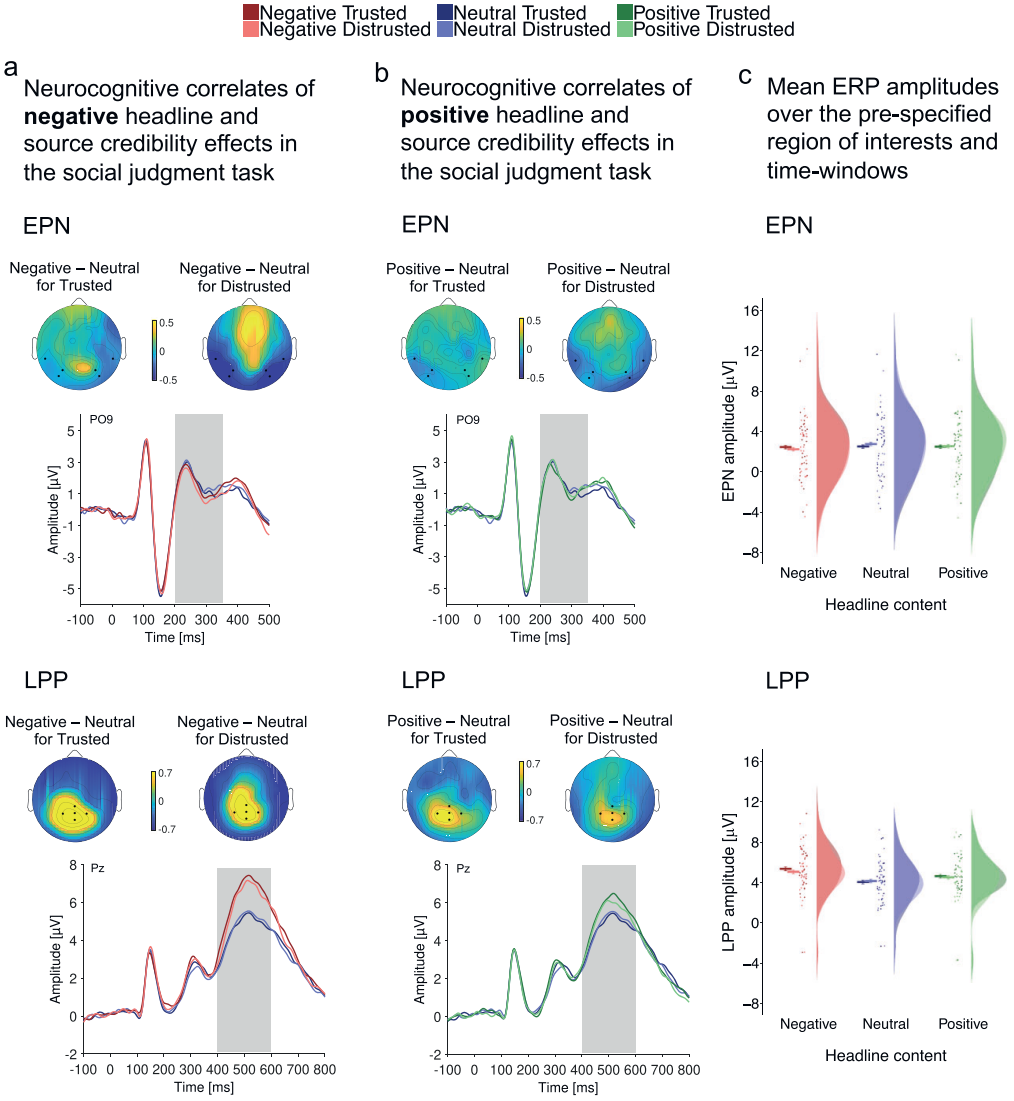

Fig. 3. In Phase 2, the EEG was acquired while social judgments were performed to investigate the neurocognitive correlates of emotional news and source credibility effects. (a) ERP results for persons related to negative headline content reveal that reflexive emotional processing in the EPN (200–350 ms) was affected by headline content. Evaluative processing in the LPP (400–600 ms) was enhanced for negative headlines from trusted as well as distrusted sources. (b) For persons related to positive headlines no EPN (200–350 ms) modulation was observed, and the LPP (400–600 ms) was enhanced for positive headlines from trusted and distrusted sources. In (a) and (b), grand average ERPs are shown for the EPN at electrode sites PO9 and for the LPP at Pz, and scalp distributions show the effects as differences between conditions in the respective time windows shaded in gray. (c) Mean ERP amplitude sizes are shown for the pre-specified ROIs and time window of the EPN and LPP. Raincloud plots (Allen *et al.*, 2019) show means and 95% CIs calculated with the 'summarySEwithin' function (Morey, 2008) on single trial data, and points, boxplots and distributions for data aggregated by subject.

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

	EPN				LPP			
Contrast	*b*	SE	*t*	*P*	*b*	SE	*t*	*P*
Trusted: Neg *vs* Neu	−0.08	0.15	−0.52	0.79	1.31	0.20	6.62	<0.001
Distrusted: Neg *vs* Neu	−0.50	0.15	−3.30	0.007	0.95	0.20	4.79	<0.001
Trusted: Pos *vs* Neu	−0.04	0.16	−0.27	0.786	0.60	0.19	3.20	0.031
Distrusted: Pos *vs* Neu	−0.18	0.16	−1.16	0.501	0.39	0.19	2.07	0.043

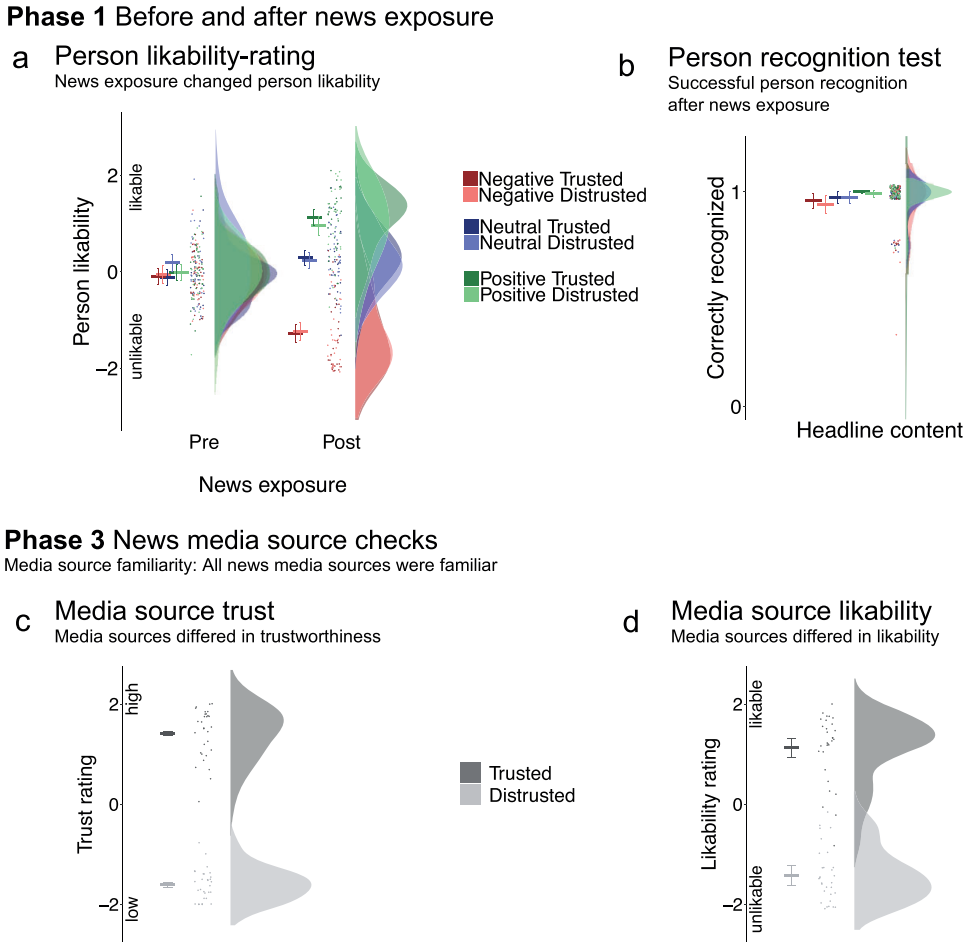

Fig. 4. In Phase 1, pre- and post-exposure person likability ratings and a post-exposure person recognition test served as manipulation checks for the news exposure. (a) Persons were liked or disliked depending on the associated headline content, unaffected by source credibility. (b) Persons were successfully recognized equally across conditions. In Phase 3, news media source checks confirmed that all sources were familiar, and that they were differentiated in (c) trustworthiness and (d) likability. Raincloud plots (Allen *et al.*, 2019) show means and 95% CIs calculated with the 'summarySEwithin' function (Morey, 2008) on single trial data, with points and distributions for data aggregated by subject.

clear and robust emotion effects were found only for distrusted sources. Future evidence should reveal additional evidence on the scope and limits of this effect. We speculate that this influence specifically of negative (but not positive) social-emotional information from distrusted sources may explain in part the popularity and success of (media) sources of questionable credibility: Untrustworthy negative social information may induce even positive states of enhanced arousal or excitation (cf. Menninghaus *et al.*, 2017), increasing the impact of negative information (cf. Kahneman and Tversky, 1979; Zillmann, 2008). Indirect evidence for a possible compounded effect of source and headline comes from research demonstrating that arousal induced by irrelevant contexts (e.g. vocal affect) can change the subsequent emotional evaluation of neutral words, an instance of evaluative conditioning (Schirmer, 2010). Taken together, we conclude that low levels of perceived credibility may, if anything, even enhance the early reception of negative headlines. As discussed above, this may be due to pleasant states of arousal associated with untrustworthy negative information (gossip) or due to a form of evaluative learning resulting in negative affect.

The trend for an EPN modulation is unlikely to be affected by the differences in perceptual salience of the different source conditions because the faces were presented in isolation during social judgment.

The present effects were observed even though participants clearly distinguished between trusted and distrusted sources, as reflected in different measures. First, the perceived credibility of the news sources was determined in a separate rating study, which was confirmed by the participants of the present study, and early emotional responses in the EPN were induced by the logos of media sources judged as untrustworthy relative to trustworthy sources (Phase 3, see Supplementary Data page 13). Please note the EPN elicited by logos is likely biased by the real-life differences in perceptual salience (e.g. red *vs* blue).

Third, active eye movements in an additional manipulation check study demonstrate that the media sources of the headlines are actively acknowledged during news exposure. Finally, we found that even the blurred website layouts without logos provide reliable cues of the source and its credibility. We are therefore confident that the credibility of media sources was successfully manipulated and noticed by the participants.

The pattern of results is in contrast to our theoretical predictions, assuming that fast reflexive processes are mainly based on the emotional contents of the headlines, whereas slower, more controlled evaluations reflected in the LPP component and the actual judgments are modulated by source credibility, putting emotional information of questionable credibility into perspective. In contrast, our findings are in line with recent evidence of strong emotion effects of untrustworthy affective person-related information. In a related study we manipulated the trustworthiness of person-related information with verbal markers such as 'supposedly' and 'people assume' (e.g. 'He allegedly bullied his trainee'; Baum *et al.*, 2018). Verbal qualifiers have an important communicative and legal function to indicate that the information might not be truthful. Just like in the present study, while participants understood the questionable veracity, person judgments and evaluative brain responses were determined by the emotional information independent of the verbally marked trustworthiness. The similarity of the findings may suggest a general mechanism.

The use of a controlled experimental design with a systematic manipulation of source credibility offers full control of confounding factors such as visual differences between faces, but it also differs in many ways from natural situations. However, here we presented existing and well-known media sources that are stored in long-term memory, including their perceived credibility. This should have even strengthened credibility effects. As in real-life situations when confronted with emotional headlines containing social information, participants in our experiment were not instructed to actively suppress the emotional content or to contemplate about the credibility of the source, but were free to consider source credibility to put their judgments into perspective. In the main task, we asked participants to repeatedly judge the person, which may have induced a strong focus on the news contents and could have them distracted from the source. However, post hoc tests including task repetition as a covariate and tests including only first judgments revealed the same pattern of results. This renders a strong bias toward social judgments and distraction away from the sources due to task repetitions unlikely. We can additionally show with eye-tracking that the source of the information is actively acknowledged during news exposure. We would also like to note that judging others based on visual appearance or minimal person-related information seems to be a natural tendency—we spontaneously form impressions about others and draw inferences about their character from minimal information (Foster, 2004; Todorov *et al.*, 2007; Bliss-Moreau *et al.*, 2008; Uhlmann *et al.*, 2015). We therefore have no reason to assume that the results are due to the experimental situation. Indeed, in a short interview after the experiment (available from 29 participants), 27 expressed no doubt about the authenticity of the media reports. Taken together, our findings complement recent online studies on how true, misleading or false information spreads and how news and its sources are evaluated (e.g. Brady *et al.*, 2017; Pennycook and Rand, 2018; Vosoughi *et al.*, 2018) by providing experimental insight into the precise neurocognitive mechanisms that underlie such behavior.

The current study was explicitly designed so that influences of the visual appearance of the faces were controlled for by careful counterbalancing. Facial trustworthiness can however influence person perception and memory (Wendt *et al.*, 2019; Lischke *et al.*, 2018; Weymar *et al.*, 2019), and thus, it would be interesting to investigate how facial-appearance-based information, such as trustworthiness, may modulate the effects of emotional information and its credibility. First evidence suggests independence of emotional information and facial appearance (Mattarozzi *et al.*, 2014; Eiserbeck and Abdel Rahman, 2020).

We conclude that the influence of source credibility on the effects of emotional contents of news headlines is remarkably weak. It is conceivable that source credibility did not qualify judgments because participants merely remembered the emotional content of the news but not the source (cf. Johnson *et al.*, 1993; Yonelinas, 2002) or that they deliberately or unintentionally ignored the credibility of the source. This distinction cannot be made based on the current results and may be targeted in future studies. Future studies may identify the circumstances under which the influence of source credibility is strengthened. This may for example depend on how salient the source is and how clearly it is represented in memory and contextually available. Future research may further target emotion regulation (Maroney and Gross, 2014; Gross, 2015) and enhanced awareness about the consequences of potentially misleading information from sources of questionable credibility as a protection against biased social judgments.

Author contributions

J.B. and R.A.R designed research and wrote the paper; J.B. performed research and analyzed data.

Acknowledgements

We thank Guido Kiecker for technical support and Cornelius Braun, Alexander Enge, Anna Faschinger, Kirsten Stark and Martin Zielinski for assisting in stimulus preparation and/or data acquisition.

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 are available online (osf.io/dwesx/).

Conflict of interest

None declared.

Supplementary data

Supplementary data are available at *SCAN* online.

References

Abdel Rahman, R. (2011). Facing good and evil: early brain signatures of affective biographical knowledge in face recognition. *Emotion*, *11*(6), 1397–405.

Abdel Rahman, R., Sommer, W. (2012). Knowledge scale effects in face recognition: an electrophysiological investigation. *Cognitive, Affective & Behavioral Neuroscience*, *12*(1), 161–74.

Allen, M., Poggiali, D., Whitaker, K., Marshall, T.R., Kievit, R.A. (2019). Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome Open Research*, 4, 63.

Aral, S., Eckles, D. (2019). Protecting elections from social media manipulation. *Science*, *365*(6456), 858–61.

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

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015b). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Baum, J., Rabovsky, M., Rose, S.B., Abdel Rahman, R. (2018). Clear judgments based on unclear evidence: person evaluation is strongly influenced by untrustworthy gossip. *Emotion*, *20*(2), 248–60.

Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300.

Berg, P., Scherg, M. (1991). Dipole models of eye movements and blinks. *Electroencephalography and Clinical Neurophysiology*, *79*, 36–44.

Blechert, J., Sheppes, G., Di Tella, C., Williams, H., Gross, J.J. (2012). See what you think: reappraisal modulates behavioral and neural responses to social stimuli. *Psychological Science*, *23*(4), 346–53.

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), 479–93.

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. *Proceedings of the National Academy of Sciences*, *114*(28), 7313–8.

Christensen, R.H.B. (2019). Ordinal – regression models for ordinal data. R package version 2019.12-10. Available: https://CRAN.R-project.org/package=ordinal.

Cunningham, W.A., Zelazo, P.D. (2007). Attitudes and evaluations: a social cognitive neuroscience perspective. *Trends in Cognitive Sciences*, *11*(3), 97–104.

De Houwer, J., Van Dessel, P., Moran, T. (2020). Attitudes beyond associations: on the role of propositional representations in stimulus evaluation. In: Bertram, Gawronski, editor. *Advances in Experimental Social Psychology*, Vol. 61, Academic Press, 127–83, ISSN 0065-2601, ISBN 9780128203729.

Delorme, A., Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.

Eiserbeck, A., Abdel Rahman, R. (2020). Visual consciousness of faces in the attentional blink: knowledge-based effects of trustworthiness dominate over appearance-based impressions. *Consciousness and Cognition*, *83*, 102977.

Ellsworth, P.C., Scherer, K.R. (2003). Appraisal processes in emotion. In: Davidson, R.J., Scherer, K.R., Goldsmith, H.H., editors. *Series in Affective Science*. New York, NY: Oxford University Press, 572–95.

Ferrari, C., Oh, D., Labbree, B.P., Todorov, A. (2020). Learning the affective value of people: more than affect-based mechanisms. *Acta Psychologica*, *203*, 103011.

Foster, E.K. (2004). Research on gossip: taxonomy, methods, and future directions. *Review of General Psychology*, *8*(2), 78–99.

Frömer, R., Maier, M., Abdel Rahman, R. (2018). Group-level EEG-processing pipeline for flexible single trial-based analyses including linear mixed models. *Frontiers in Neuroscience*, *12*, 970.

Galli, G., Feurra, M., Viggiano, M.P. (2006). "Did you see him in the newspaper?" Electrophysiological correlates of context and valence in face processing. *Brain Research*, *1119*(1), 190–202.

Gawronski, B., Bodenhausen, G.V. (2006). Associative and propositional processes in evaluation: an integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692–731.

Gross, J.J. (2015). Emotion regulation: current status and future prospects. *Psychological Inquiry*, *26*(1), 1–26.

Herbert, C., Pauli, P., Herbert, B.M. (2011). Self-reference modulates the processing of emotional stimuli in the absence of explicit self-referential appraisal instructions. *Social Cognitive and Affective Neuroscience*, *6*(5), 653–61.

Herbert, C., Sfärlea, A., Blumenthal, T. (2013). Your emotion or mine: labeling feelings alters emotional face perception—an ERP study on automatic and intentional affect labeling. *Frontiers in Human Neuroscience*, *7*.

Johnson, M.K., Hashtroudi, S., Lindsay, D.S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3–28.

Junghöfer, M., Bradley, M.M., Elbert, T.R., Lang, P.J. (2001). Fleeting images: a new look at early emotion discrimination. *Psychophysiology*, *38*(2), 175–8.

Junghöfer, M., Rehbein, M.A., Maitzen, J., Schindler, S., Kissler, J. (2016). An evil face? Verbal evaluative multi-CS conditioning enhances face-evoked mid-latency magnetoencephalographic responses. *Social Cognitive and Affective Neuroscience*, *76*(3), nsw179–11.

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American Psychologist*, *58*(9), 697–720.

Kahneman, D., Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, *47*(2), 263.

Kissler, J., Herbert, C., Peyk, P., Junghöfer, M. (2007). Buzzwords early cortical responses to emotional words during reading. *Psychological Science*, *18*(6), 475–80.

Kissler, J., Strehlow, J. (2017). Something always sticks? How emotional language modulates neural processes involved in face encoding and recognition memory. *Poznan Studies in Contemporary Linguistics*, *53*(1), 109–32.

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B. (2017). lmerTestPackage: tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26.

Lazer, D.M.J., Baum, M.A., Benkler, Y., et al. (2018). The science of fake news. *Science*, *359*(6380), 1094–6.

Lenth, R. (2020). emmeans: estimated marginal means, aka least-squares means. R package version 1.4.6. Available: https://CRAN.R-project.org/package=emmeans.

Lewandowsky, S., Ecker, U.K.H., Seifert, C.M., Schwarz, N., Cook, J. (2012). Misinformation and its correction: continued influence and successful debiasing. *Psychological Science in the Public Interest*, *13*(3), 106–31.

Lieberman, M.D. (2007). Social cognitive neuroscience: a review of core processes. *Annual Review of Psychology*, *58*(1), 259–89.

Lischke, A., Junge, M., Hamm, A.O., Weymar, M. (2018). Enhanced processing of untrustworthiness in natural faces with neutral expressions. *Emotion*, *18*(2), 181–9.

Luck, S.J. (2014). *An Introduction to the Event-related Potential Technique*. Cambridge, MA: MIT Press.

Luo, Q.L., Wang, H.L., Dzhelyova, M., Huang, P., Mo, L. (2016). Effect of affective personality information on face processing: evidence from ERPs. *Frontiers in Psychology*, *7*, 1397.

Maroney, T.A., Gross, J.J. (2014). The ideal of the dispassionate judge: an emotion regulation perspective. *Emotion Review*, *6*(2), 142–51.

Menninghaus, W., Wagner, V., Hanich, J., Wassiliwizky, E., Jacobsen, T., Koelsch, S. (2017). The distancing-embracing model of the enjoyment of negative emotions in art reception. *Behavioral and Brain Sciences*, *40*, 26.

Mattarozzi, K., Todorov, A., Codispoti, M. (2014). Memory for faces: the effect of facial appearance and the context in which the face is encountered. *Psychological Research*, *79*(2), 308–17.

Metzger, M.J., Flanagin, A.J. (2013). Credibility and trust of information in online environments: the use of cognitive heuristics. *Journal of Pragmatics*, *59*, 210–20.

Morey, R.D. (2008). Confidence intervals from normalized data: a correction to Cousineau (2005). *Tutorials in Qualitative Methods for Psychology*, 4, 61–4.

Pennycook, G., Rand, D.G. (2018). Lazy, not biased: susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, *188*, 39–50.

Pennycook, G., Rand, D.G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences*, *116*(7), 2521–6.

Rellecke, J., Sommer, W., Schacht, A. (2012). Does processing of emotional facial expressions depend on intention? Time-resolved evidence from event-related brain potentials. *Biological Psychology*, *90*(1), 23–32.

Sabatinelli, D., Keil, A., Frank, D.W., Lang, P.J. (2013). Emotional perception: correspondence of early and late event-related potentials with cortical and subcortical functional MRI. *Biological Psychology*, *92*(3), 513–9.

Schirmer, A. (2010). Mark my words: tone of voice changes affective word representations in memory. *PLoS One*, *5*(2), e9080.

Schad, D.J., Vasishth, S., Hohenstein, S., Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models_ a tutorial. *Journal of Memory and Language*, *110*, 104038.

Schacht, A., Sommer, W. (2009a). Emotions in word and face processing: early and late cortical responses. *Brain and Cognition*, *69*(3), 538–50.

Schacht, A., Sommer, W. (2009b). Time course and task dependence of emotion effects in word processing. *Cognitive, Affective & Behavioral Neuroscience*, *9*(1), 28–43.

Scherer, K.R. (2001). Appraisal considered as a process of multilevel sequential checking. In: Scherer, K.R., Schorr, A., Johnstone, T., editors. *Series in Affective Science*. New York, NY: Oxford University Press, 92–120.

Schindler, S., Vormbrock, R., Kissler, J. (2019). Emotion in context: how sender predictability and identity affect processing of words as imminent personality feedback. *Frontiers in Psychology*, *10*, 327.

Schupp, H.T., Junghöfer, M., Weike, A.I., Hamm, A.O. (2003). Emotional facilitation of sensory processing in the visual cortex. *Psychological Science*, *14*(1), 7–13.

Schupp, H.T., Öhman, A., Junghöfer, M., Weike, A.I., Stockburger, J., Hamm, A.O. (2004). The facilitated processing of threatening faces: an ERP analysis. *Emotion*, *4*(2), 189–200.

Strack, F., Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*(3), 220–47.

Suess, F., Rabovsky, M., Abdel Rahman, R. (2015). Perceiving emotions in neutral faces: expression processing is biased by affective person knowledge. *Social Cognitive and Affective Neuroscience*, *10*(4), 531–6.

Todorov, A., Gobbini, M.I., Evans, K.K., Haxby, J.V. (2007). Spontaneous retrieval of affective person knowledge in face perception. *Neuropsychologia*, *45*(1), 163–73.

Uhlmann, E.L., Pizarro, D.A., Diermeier, D. (2015). A person-centered approach to moral judgment. *Perspectives on Psychological Science*, *10*(1), 72–81.

Vosoughi, S., Roy, D., Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–51.

Wendt, J., Weymar, M., Junge, M., Hamm, A.O., Lischke, A. (2019). Heartfelt memories: cardiac vagal tone correlates with increased memory for untrustworthy faces. *Emotion*, *19*(1), 178–82.

Weymar, M., Ventura-Bort, C., Wendt, J., Lischke, A. (2019). Behavioral and neural evidence of enhanced long-term memory for untrustworthy faces. *Scientific Reports*, *9*(1), 19217–18.

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.

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. *Frontiers in Psychology*, *7*(1368), 1446.

Yonelinas, A.P. (2002). The nature of recollection and familiarity: a review of 30 years of research. *Journal of Memory and Language*, *46*(3), 441–517.

Zillmann, D. (2008). *Excitation Transfer Theory*, Vol. 89, Chichester, UK: John Wiley & Sons, Ltd.

NeuroImage 244 (2021) 118572

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

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

ARTICLE INFO

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

ABSTRACT

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

* Corresponding authors at: Humboldt-Universität zu Berlin, Institut für Psychologie, Unter den Linden 6, 10099 Berlin, Germany.
E-mail addresses: julia.baum@hu-berlin.de (J. Baum), rasha.abdel.rahman@hu-berlin.de (R. Abdel Rahman).

https://doi.org/10.1016/j.neuroimage.2021.118572.
Received 26 May 2021; Received in revised form 10 August 2021; Accepted 7 September 2021
Available online 8 September 2021.

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; Kissler 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 Flanagin, 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; Kahneman, 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. Sabatinelli 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; Rellecke 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 eyesight, 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., 2010; 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 × 3.5 cm, viewing distance 70 cm). Eight

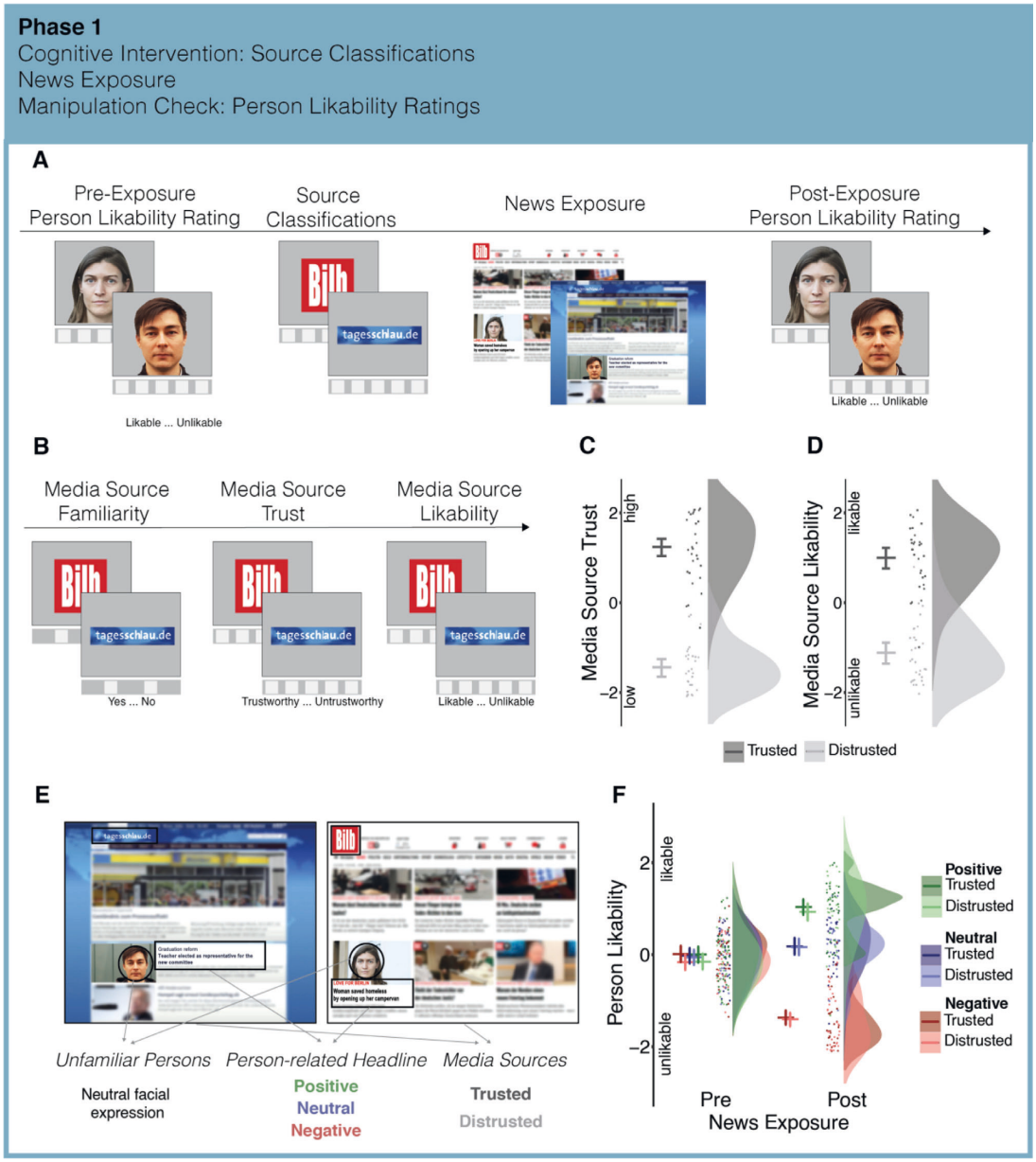

Fig. 1. Overview of the study design with two phases (Phase 1 is shown here) and key experimental manipulations. **A.** Depicted is the procedure of Phase 1 with the cognitive intervention in form of source classifications, the news exposure, and framing pre- and post-exposure person likability ratings 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 teenager 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 increments; 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=0.39, 95%-CI [.04, 1.57], t = 2.07, p=.046; SI-Table S18, S19), this direction was unexpected and could be explained by participants giving a rather controlled and cognitive answer, possibly related to or confounded with the previously rated trustworthiness, possibly reflecting a tendency for giving social desirable answers. As expected, the EPN was enhanced for distrusted compared to trusted sources (b = 1.08, SE=0.41, 95%-CI [.27, 1.89], t = 2.07, p=.026; SI-Table S20 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 70minutes 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, PO9, 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, Kuznetsova et al., 2017). For each dependent variable, the model was specified with fixed effects for the experimental factors *headline content* (negative, positive, 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 (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, $F(1,29)=0$, $p = 1$).

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 <|.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

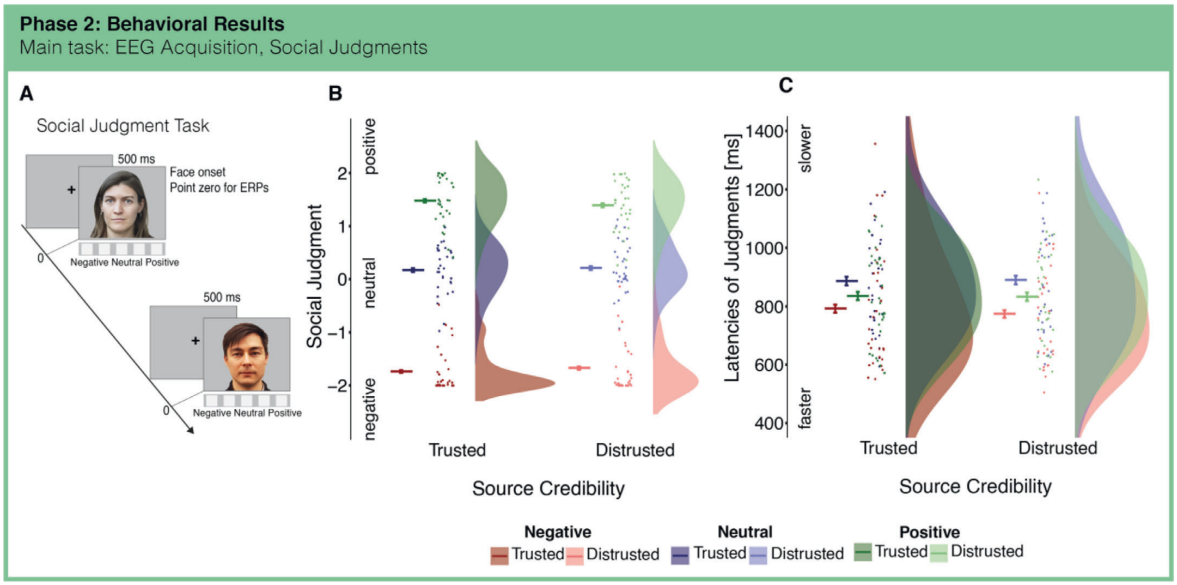

Fig. 2. Behavioral results in the social judgment task. In Phase 2 the social judgment was performed as main task to investigate the effects of source credibility and headline content. **A.** Participants judged persons presented in isolation based on the information from Phase 1. Faces were presented 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 for negative 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.

	Social Judgments				Latencies of Judgments [−1000/latency(ms)]			
Coefficient	*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 (−0.02 – 0.04)	0.01	0.48	0.634
Negative Headline Content (Neg vs. Neu)	−1.89 (−2.13 – −1.66)	0.12	−16.00	**<0.001**	−0.16 (−0.22 – −0.09)	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 (−0.05 – 0.08)	0.03	0.41	0.685
Positive Headline Content (Pos vs. Neu)	1.24 (1.04 – 1.45)	0.11	11.82	**<0.001**	−0.07 (−0.12 – −0.01)	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 (−0.08 – 0.06)	0.03	−0.31	0.759
Model Formula	Judgment ~ Headline Content * Source Credibility + (*S*+Neg+*S**Neg+Pos+*S**Pos | subject) + (*S*+Neg+*S**Neg+Pos+*S**Pos | face)				Latency ~ Headline Content * Source Credibility + (*S*+Neg+*S**Neg+Pos+*S**Pos | subject) + (*S*+Neg+*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.

	Social Judgments				Latencies of Judgments [−1000/latency(ms)]			
Contrast	*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.

	Overall Judgments				Confidence in Judgments			
Coefficient	*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 (1.15 – 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 (−8.94 – 8.75)	4.51	−0.02	0.984
Model Formula	Judgment ~ Headline Content * Source Credibility + (Neg+Pos || subject) + (*S*+Neg+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.

	Overall Judgments				Confidence in Judgments			
Contrast	*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.

	EPN				LPP			
Coefficient	*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*+Neg+*S**Neg+Pos+*S**Pos || face)				LPP ~ Headline Content * Source Credibility + (*S*+Neg+Pos+*S**Pos || subject) + (Neg +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.

	EPN				LPP			
Contrast	*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

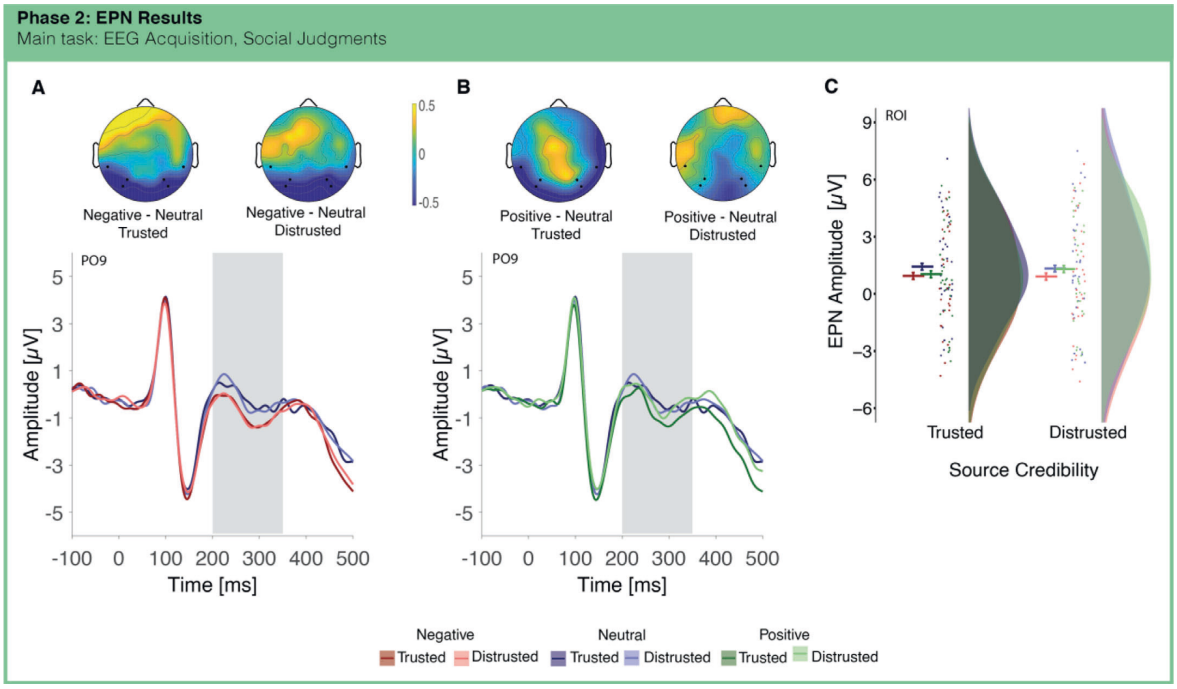

Fig. 3. Early posterior negativity (EPN) modulations by headline content and source credibility effects. In Phase 2, the EEG was acquired in the social judgment task when faces were presented in isolation. We investigated neurocognitive correlates of fast and reflexive emotional processing in the EPN (200–350 ms). **A.** The EPN 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 EPN was enhanced compared to neutral only for trusted sources, but not for distrusted sources. **A.–B.** Grand average ERPs are shown at electrode site PO9 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 *summarySEwithin* function (Morey, 2008) on single trial data, and points, boxplots, and distributions for data aggregated by subject.

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 Flanagin, 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

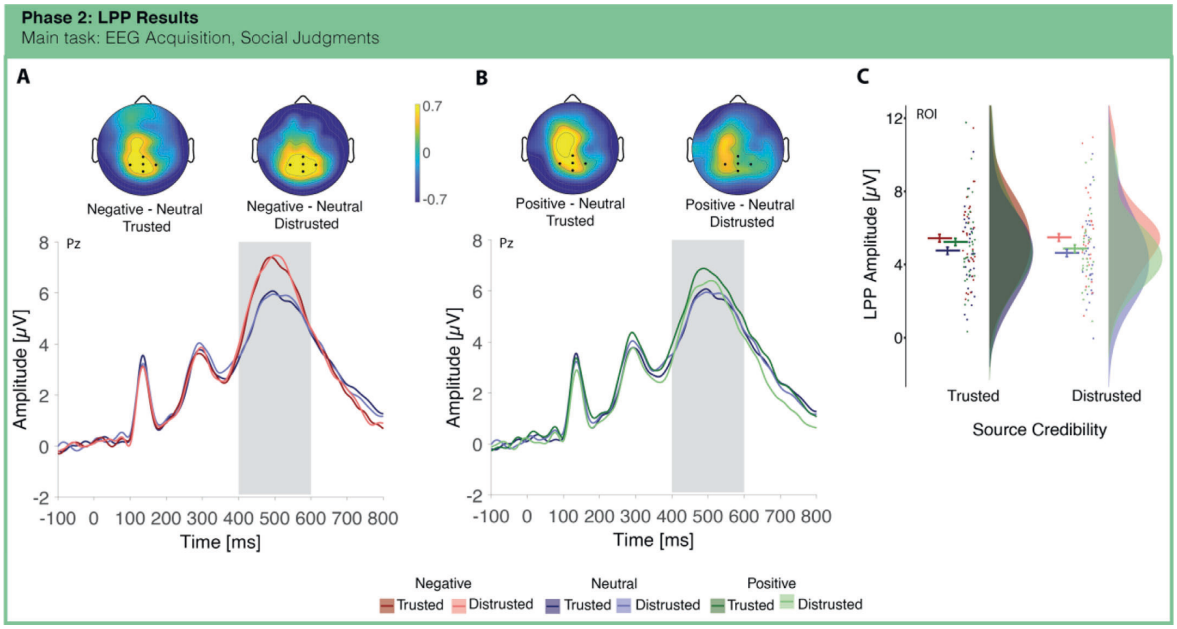

Fig. 4. Late positive potential (LPP) modulations by headline content and source credibility effects. In Phase 2, the EEG was acquired in the social judgment task when faces were presented in isolation. We investigated neurocognitive correlates of slow and evaluative processing in the LPP (400–600 ms). **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.

	N170			
Coefficient	*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
Model Formula	N170 ~ Headline Content * Source Credibility + (Pos + *S**Neg || subject) + (Neg +Pos+*S**Neg+*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.

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-

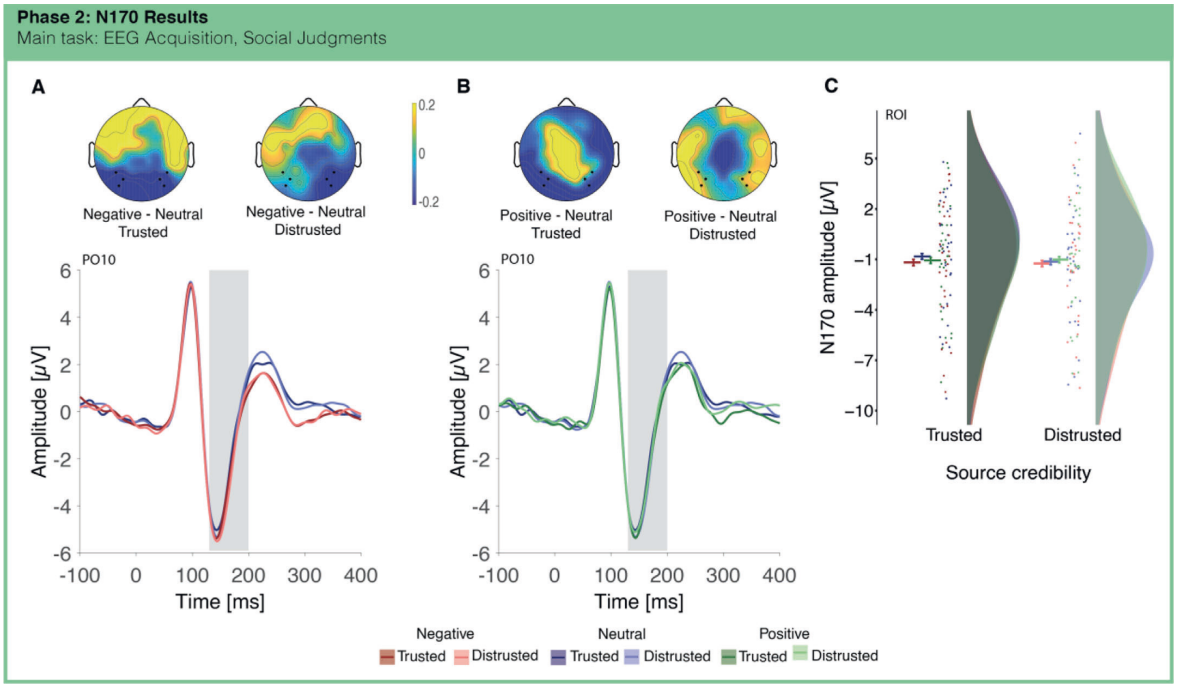

Fig. 5. N170 modulations by headline content and source credibility effects. In Phase 2, the EEG was acquired in the social judgment task when faces were presented in isolation. We investigated neurocognitive correlates of visual face processing in the N170 (130–200 ms). **A.** The N170 was enhanced 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 *summarySEwithin* 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 end of 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 $\langle$0.015 for *Trusted*, and ps $\rangle$ 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* × *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 17ff). Whereas the influence of associated negative headlines was not modulated by source credibility assignments (*Source Credibility* × *Negative Headline Content* all ps > 0.3, see SI pages 17ff).

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., Kozyreva et al., 2020; Lorenz-Spreen et al., 2020; Pennycook et al., 2020; Pennycook et al., 2021; Roozenbeek 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-

ciplines that investigate this topic from different angles. Such strategies may include educating about the impact of social-emotional misinformation on the individual level and employing emotion regulation, for instance by deliberate ignorance or reappraisal of emotional contents (Hertwig and Engel, 2016; Gross, 2015; Maroney and Gross, 2014).

3.2. Summary and conclusion

Together, our findings show important similarities in how positive and negative social-emotional headlines affect social judgments regardless of source credibility and the current attempt to increase the minding of news trustworthiness. Additionally, our findings also tentatively suggest differential effects for positive and negative contents to a certain degree, indicating that this bias is more susceptible to our intervention when the headline contents are positive, whereas it is particularly difficult to overcome for negative headlines. This might be related to the valence as such or, to the emotional intensity of the headlines irrespective of valence. A distinction cannot be drawn based on the current data. Additional analyses on ERPs suggest that valence might play an independent role, as differences between ERPs in response to positive and negative headlines were only found for distrusted sources (SI page 10). Moreover, preratings of the arousal of the headlines show that positive and negative headlines don't differ from each other statistically (SI page 11). In general, negative contents may have higher social informational value than positive information, as stronger effects of stimuli with negative compared to positive valence have frequently been found (Abdel Rahman, 2011; Anderson et al., 2011; Skowronski and Carlston, 1987; Suess et al., 2014). Thus, negative information may be processed under high priority as a protection against the potential threat an allegedly dangerous person could pose. This preferential processing seems to come at the cost of critical thinking, and recent evidence supports this by showing that negative emotional narrative contexts may reduce critical semantic analyses (Aristei et al., 2021). Furthermore, there is some evidence that meaningful contexts can modulate specifically the processing and impact of positive stimuli. For example, taking positive words out of isolation and into social-communicative situations has been shown to enhance their emotion effects in ERPs (Rohr and Abdel Rahman, 2015).

In conclusion, the current evidence suggests that negative social-emotional misinformation has prevailing consequences despite increased attention to source credibility. This abates the hope for simple solutions to the misinformation problem, especially because negative contents are often favored in online media, where attention is a scarce and economically valuable resource. As a result, consumers are confronted with contents that evoke negative emotions such as moral outrage at a high rate, as these contents can powerfully capture consumers' attention and are more readily shared (e.g., Brady et al., 2017, 2020; Crockett, 2017; Soroka et al., 2019; Schöne et al., 2021; Van Bavel et al., 2020). Conversely, this may even give an advantage to misinformation, as fear- and disgust-evoking false news have been found to travel faster and further in social networks than true news that were more characterized by anticipation or joy (Vosoughi et al., 2018). The current insights into the cognitive mechanisms of responses to misinformation in mind and brain suggest that emotional contents and how they can be put in perspective are key targets for future strategies to avoid false news' potentially detrimental effects.

Declaration of Competing Interest

The authors declare no conflicts of interest with respect to authorship or the publication of this article.

Author's contribution

J. Baum and R. Abdel Rahman designed research and wrote the paper; J. Baum performed research and analyzed data.

Acknowledgements

We thank Guido Kiecker for technical support and Friedrich Eiserbeck, Alexander Enge, 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 (6), 1397–1405. doi:10.1037/a0024717.

Allen, M., Poggiali, D., Whitaker, K., Marshall, T.R., Kievit, 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 (6036), 1446–1448.

Aristei, S., Knoop, C.A., Lubrich, O., Nehrlich, 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., Ensenberg, N., Hassin, R.R., 2017. The inherently contextualized 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. Stat. 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/nsaa164.

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/3zr24.

Baum, J., Rabovsky, M., Rose, S.B., Abdel Rahman, R., 2018. Clear judgments based on unclear evidence: person evaluation is strongly influenced by untrustworthy gossip. Emotion 20 (2), 248–260. doi:10.1037/emo0000545.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Statis. Soc. Series B (Methodological) 57 (1), 289–300 10.2307/2346101?refreqid=search-gateway:025a1890f1052cf99b3585fdaa4b55fd.

Berg, P., Scherg, M., 1991. Dipole models of eye movements and blinks. Electroencephalography Clin. Neurophysiol. 79 (1), 36–44. doi:10.1016/0013-4694(91)90154-V.

Blechert, J., Sheppes, G., Di Tella, C., Williams, H., Gross, J.J., 2012. See what you think: reappraisal modulates behavioral and neural responses to social stimuli. Psychol. Sci. 23 (4), 346–353. doi:10.1177/0956797612438559.

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.2006.12.005.

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.

Dias, N., Pennycook, G., Rand, D.G., 2020. Emphasizing publishers does not effectively reduce susceptibility to misinformation on social media. Harvard Kennedy School Misinformat. Rev. 1–12.
Ebner, N.C., Riediger, M., Lindenberger, U., 2010. FACES—A database of facial expressions in young, middle-aged, and older women and men: development and validation. Behav. Res. Methods 42 (1), 351–362. doi:10.3758/BRM.42.1.351.
Eiserbeck, A., Abdel Rahman, R., 2020. Visual consciousness of faces in the attentional blink: knowledge-based effects of trustworthiness dominate over appearance-based impressions. Conscious. Cogn. 83, 102977. doi:10.1016/j.concog.2020.102977.
Eiserbeck, A., Enge, A., Rabovsky, M., Abdel Rahman, R., 2021. Distrust before first sight: Knowledge- and appearancebased effects of trustworthiness on the visual consciousness of faces. BioRxiv doi:10.1101/2021.02.24.432562.
Ellsworth, P.C., Scherer, K.R, 2003. Appraisal processes in emotion. In: Davidson, R.J., Scherer, K.R., Goldsmith, H.H. (Eds.), Series in Affective Science. Oxford University Press, New York, NY, US, pp. 572–595.
Ensenberg, N.S., Perry, A., Aviezer, H., 2017. Are you looking at me? Mu suppression modulation by facial expression direction. Cognit., Affect. Behav. Neurosci. 17 (1), 174–184. doi:10.3758/s13415-016-0470-z.
FeldmanHall, O., Shenhav, A., 2019. Resolving uncertainty in a social world. Nature Human Behav. 11 (5), 419. doi:10.1038/s41562-019-0590-x.
Ferrari, C., Oh, D., Labbree, B.P., Todorov, A., 2020. Learning the affective value of people: more than affect-based mechanisms. Acta Psychol. (Amst) 203, 103011. doi:10.1016/j.actpsy.2020.103011.
Frömer, R., Maier, M., Abdel Rahman, R., 2018. Group-Level EEG-processing pipeline for flexible single trial-based analyses including linear mixed models. Front Neurosci 12, 970. doi:10.3389/fnins.2018.00048.
Frömer, R., Nassar, M.R., Bruckner, R., Stürmer, B., Sommer, W., Yeung, N., 2021. Response-based outcome predictions and confidence regulate feedback processing and learning. Elife 10, 1–29. doi:10.7554/ELIFE.62825.
Galli, G., Feurra, M., Viggiano, M.P., 2006. Did you see him in the newspaper?" Electrophysiological correlates of context and valence in face processing. Brain Res. 1119 (1), 190–202. doi:10.1016/j.brainres.2006.08.076.
Gawronski, B., Bodenhausen, G.V., 2006. Associative and propositional processes in evaluation: an integrative review of implicit and explicit attitude change. Psychol. Bull. 132 (5), 692–731. doi:10.1037/0033-2909.132.5.692.
Giménez-Fernández, T., Fernández-Folgueiras, U., Fondevila, S., Méndez-Bértolo, C., García-Rubio, M.J., Hernández-Lorca, M., Kessel, D., Carretié, L., 2021. Enhanced N170 to outgroup faces: perceptual novelty or prejudice? Soc. Neurosci. 16 (3), 252–264. doi:10.1080/17470919.2021.1889658.
Giménez-Fernández, T., Kessel, D., Fernández-Folgueiras, U., Fondevila, S., Méndez-Bértolo, C., Aceves, N., García-Rubio, M.J., Carretié, L., 2020. Prejudice drives exogenous attention to outgroups. Soc. Cogn. Affect. Neurosci. 15 (6), 615–624. doi:10.1093/scan/nsaa087.
Gordon, A., Quadflieg, S., Brooks, J.C.W., Ecker, U.K.H., Lewandowsky, S., 2019. Keeping track of 'alternative facts': the neural correlates of processing misinformation corrections. Neuroimage 193, 46–56. doi:10.1016/j.neuroimage.2019.03.014.
Gross, J.J., 2015. Emotion Regulation: current Status and Future Prospects. Psychol. Inq 26 (1), 1–26. doi:10.1080/1047840X.2014.940781.
Herbert, C., Pauli, P., Herbert, B.M., 2011. Self-reference modulates the processing of emotional stimuli in the absence of explicit self-referential appraisal instructions. Soc. Cogn. Affect. Neurosci. 6 (5), 653–661. doi:10.1093/scan/nsq082.
Herbert, C., Sfärlea, A., Blumenthal, T., 2013. Your emotion or mine: labeling feelings alters emotional face perception—An ERP study on automatic and intentional affect labeling. Front. Hum. Neurosci. 7. doi:10.3389/fnhum.2013.00378.
Hertwig, R., Engel, C., 2016. Homo Ignorans: deliberately Choosing Not to Know. Perspect. Psychol. Sci. 11 (3), 359–372. doi:10.1177/1745691616635594.
Hertwig, R., Grüne-Yanoff, T., 2017. Nudging and Boosting: Steering or Empowering Good Decisions. Perspect. Psychological Sci. 12 (6), 973–986. doi:10.1177/1745691617702496.
Hinojosa, J.A., Mercado, F., Carretié, L., 2015. N170 sensitivity to facial expression: a meta-analysis. Neurosci. Biobehav. Rev. 55, 498–509. doi:10.1016/j.neubiorev.2015.06.002.
Johnson, M.K., Hashtroudi, S., Lindsay, D.S., 1993. Source monitoring. Psychol. Bull. 114 (1), 3–28. doi:10.1037//0033-2909.114.1.3.
Junghöfer, M., Bradley, M.M., Elbert, T.R., Lang, P.J., 2001. Fleeting images: a new look at early emotion discrimination. Psychophysiology 38 (2), 175–178. doi:10.1111/1469-8986.3820175.
Junghöfer, M., Rehbein, M.A., Maitzen, J., Schindler, S., Kissler, J., 2016. An evil face? Verbal evaluative multi-CS conditioning enhances face-evoked mid-latency magnetoencephalographic responses. Soc. Cogn. Affect. Neurosci. 76 (3), nsw179. doi:10.1093/scan/nsw179, –11.
Kahneman, D., 2003. A perspective on judgment and choice: mapping bounded rationality. Am. Psychol. 58 (9), 697–720. doi:10.1037/0003-066X.58.9.697.
Kissler, J., Strehlow, J., 2017. Something always sticks? how emotional language modulates neural processes involved in face encoding and recognition memory. Poznan Stud. Contemporary Linguist. 53 (1), 109–132. doi:10.1515/psicl-2017-0004.
Kozyreva, A., Lewandowsky, S., Hertwig, R., 2020. Citizens versus the internet: confronting digital challenges with cognitive tools. Psychol. Sci. Public Interest 21 (3), 103–156. doi:10.1177/1529100620946707.
Krasowski, C., Schindler, S., Bruchmann, M., Moeck, R., Straube, T., 2021. Electrophysiological responses to negative evaluative person-knowledge: effects of individual differences. Cognit., Affect. Behav. Neurosci. doi:10.3758/s13415-021-00894-w.
Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. lmerTestPackage: tests in linear mixed effects models. J. Stat. Softw. 82 (13), 1–26. doi:10.18637/jss.v082.i13.
Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H.J., Hawk, S.T., van Knippenberg, A., 2010. Presentation and validation of the radboud faces database. Cognit. Emotion 24 (8), 1377–1388. doi:10.1080/02699930903485076.
Lazer, D.M.J., Baum, M.A., Benkler, Y., Berinsky, A.J., Greenhill, K.M., Menczer, F., et al., 2018. The science of fake news: Addressing fake news requires a multidisciplinary effort. Science 359 (6380), 1094–1096. doi:10.1126/science.aao2998.
Lenth, R., (2020). emmeans: estimated marginal means, aka least-squares means. R package version 1.4.6. https://CRAN.R-project.org/package=emmeans
Lieberman, M.D., 2007. Social cognitive neuroscience: a review of core processes. Annu. Rev. Psychol. 58 (1), 259–289. doi:10.1146/annurev.psych.58.110405.085654.
Lorenz-Spreen, P., Lewandowsky, S., Sunstein, C.R., Hertwig, R., 2020. How behavioural sciences can promote truth, autonomy and democratic discourse online. Nature Human Behav. 1–8. doi:10.1038/s41562-020-0889-7.
Luck, S.J., 2014. An Introduction to the Event-Related Potential Technique. MIT press.
Luo, Q.L., Wang, H.L., Dzhelyova, M., Huang, P., Mo, L., 2016. Effect of affective personality information on face processing: evidence from ERPs. Front. Psychol. 7, 1397. doi:10.3389/fpsyg.2016.00810.
Maroney, T.A., Gross, J.J., 2014. The Ideal of the Dispassionate Judge: an Emotion Regulation Perspective. Emotion Review 6 (2), 142–151. doi:10.1177/1754073913491989.
Martel, C., Pennycook, G., Rand, D.G., 2020. Reliance on emotion promotes belief in fake news. Cognit. Res. 5 (1), 1–20. doi:10.1186/s41235-020-00252-3.
Mattarozzi, K., Todorov, A., Codispoti, M., 2014. Memory for faces: the effect of facial appearance and the context in which the face is encountered. Psychol. Res. 79 (2), 308–317. doi:10.1007/s00426-014-0554-8.
Metzger, M.J., Flanagin, A.J., 2013. Credibility and trust of information in online environments: the use of cognitive heuristics. J Pragmat 59, 210–220. doi:10.1016/j.pragma.2013.07.012.
Morey, R.D., 2008. Confidence intervals from normalized data: a correction to Cousineau (2005). Tutor. Qual. Methods Psychol. 4, 61–64.
Öhman, A., Mineka, S., 2001. Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. Psychological Rev. 108 (3), 483–522. doi:10.1037/0033-295X.108.3.483.
Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A.A., Eckles, D., Rand, D.G., 2021. Shifting attention to accuracy can reduce misinformation online. Nature (April) 592. doi:10.1038/s41586-021-03344-2.
Pennycook, G., McPhetres, J., Zhang, Y., Lu, J.G., Rand, D.G., 2020. Fighting COVID-19 Misinformation on Social Media: experimental Evidence for a Scalable Accuracy-Nudge Intervention. Psychol. Sci. 31 (7), 770–780.
Pennycook, G., Rand, D.G., 2019. Fighting misinformation on social media using crowdsourced judgments of news source quality. Proc. Natl. Acad. Sci. 116 (7), 2521–2526. doi:10.1073/pnas.1806781116.
Rabovsky, M., Stein, T., Abdel Rahman, R., 2016. Access to awareness for faces during continuous flash suppression is not modulated by affective knowledge. PLoS ONE 11 (4), e0150931. doi:10.1371/journal.pone.0150931.
Rellecke, J., Sommer, W., Schacht, A., 2012. Does processing of emotional facial expressions depend on intention? Time-resolved evidence from event-related brain potentials. Biol. Psychol. 90 (1), 23–32. doi:10.1016/j.biopsycho.2012.02.002.
Rohr, L., Abdel Rahman, R., 2015. Affective responses to emotional words are boosted in communicative situations. Neuroimage 109 (C), 273–282. doi:10.1016/j.neuroimage.2015.01.031.
Roozenbeek, J., van der Linden, S., 2019. Fake news game confers psychological resistance against online misinformation. Palgrave Communications 5 (1), 1–10. doi:10.1057/s41599-019-0279-9.
Sabatinelli, D., Keil, A., Frank, D.W., Lang, P.J., 2013. Emotional perception: correspondence of early and late event-related potentials with cortical and subcortical functional MRI. Biol. Psychol. 92 (3), 513–519. doi:10.1016/j.biopsycho.2012.04.005.
Schad, D.J., Vasishth, S., Hohenstein, S., Kliegl, R., 2020. How to capitalize on a priori contrasts in linear (mixed) models_ A tutorial. J. Mem Lang 110, 104038. doi:10.1016/j.jml.2019.104038.
Schacht, A., Sommer, W., 2009a. Emotions in word and face processing: early and late cortical responses. Brain Cogn. 69 (3), 538–550. doi:10.1016/j.bandc.2008.11.005.
Schacht, A., Sommer, W., 2009b. Time course and task dependence of emotion effects in word processing. Cognit., Affect., Behav. Neurosci. 9 (1), 28–43. doi:10.3758/CABN.9.1.28.
Scherer, K.R., 2001. Appraisal considered as a process of multilevel sequential checking. In: Scherer, K.R., Schorr, A., Johnstone, T. (Eds.), Series in Affective Science. Oxford University Press, New York, NY, US, pp. 92–120.
Schindler, S., Bruchmann, M., Steinweg, A.-L., Moeck, R., Straube, T., 2020. Attentional conditions differentially affect early, intermediate and late neural responses to fearful and neutral faces. Soc. Cogn. Affect. Neurosci. 15 (7), 765–774. doi:10.1093/scan/nsaa098.
Schindler, S., Bruchmann, M., Krasowski, C., Moeck, R., Straube, T., 2021. Charged with a crime: the neuronal signature of processing negatively evaluated faces under different attentional conditions. Psychol. Sci. doi:10.1177/0956797621996667.
Schindler, S., Vormbrock, R., Kissler, J., 2019. Emotion in Context: how Sender Predictability and Identity Affect Processing of Words as Imminent Personality Feedback. Front. Psychol. 10, 327. doi:10.3389/fpsyg.2019.00094.
Schöne, J., Parkinson, B., & Goldenberg, A. (2021, January 2). Negativity Spreads More than Positivity on Twitter after both Positive and Negative Political Situations. 10.31234/osf.io/x9e7u
Schupp, H.T., Junghöfer, M., Weike, A.I., Hamm, A.O., 2003. Emotional Facilitation of Sensory Processing in the Visual Cortex. Psychol. Sci. 14 (1), 7–13. doi:10.1111/1467-9280.01411.
Schupp, H.T., Öhman, A., Junghöfer, M., Weike, A.I., Stockburger, J., Hamm, A.O., 2004. The Facilitated Processing of Threatening Faces: an ERP Analysis. Emotion 4 (2), 189–200. doi:10.1037/1528-3542.4.2.189.

Skowronski, J.J., Carlston, 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., Gobbini, M.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, I.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-Centered 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:10.1016/j.biopsycho.2014.11.012.

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.