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-likeability 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 likeability, 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 we 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 we consider the lack of reliability when we subjectively judge persons as negative, neutral or positive based on verbal information about their immoral or moral social behavior.

How we perceive, judge, and interact with other people is strongly influenced by what we know about them. Even abstract and verbally transmitted information concerning their good or bad social behavior can affect how we judge others (Bliss-Moreau, Barrett, & Wright, 2008; Goodwin, Piazza, & Rozin, 2014), how we perceive their 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 we see their 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, while the LPP has been shown to vary with the relevance and meaning of emotional attributes in a given context, the earlier emotion-induced reflexive EPN modulations are relatively independent of task demands and the relevance of emotional contents, and 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 independent of the verbally marked trustworthiness of the information at early stages reflected in the EPN, while 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 provides 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 likeability 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, e.g. by labelling an unpleasant scene as "fictitious" vs. "real" (Mocaiber et al., 2010), or by reappraising a person's angry face with their 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 we use context information to adjust our

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 labelled 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, i.e. immoral, and positive information was about kind social behaviour, i.e. moral (*cf.* Hofmann, Wisneski, Brandt, & Skitka, 2014). Subsequent to learning a test phase was conducted in which participants 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 likeability 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 likeability 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

Methods

Sample size

For experiment 1 a multiple of 8 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, 3 measurement levels, rmANOVA within subject) 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 consists 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, e.g. *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 2 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

Means and confidence intervals for the trustworthiness rating in experiment 1

Exp.		Negative Facts	Negative Gossip
1	*M*	6.65	5.70
	95% CI	[6.25, 7.05]	[5.29, 6.10]

Note. CI = confidence interval.

Procedure

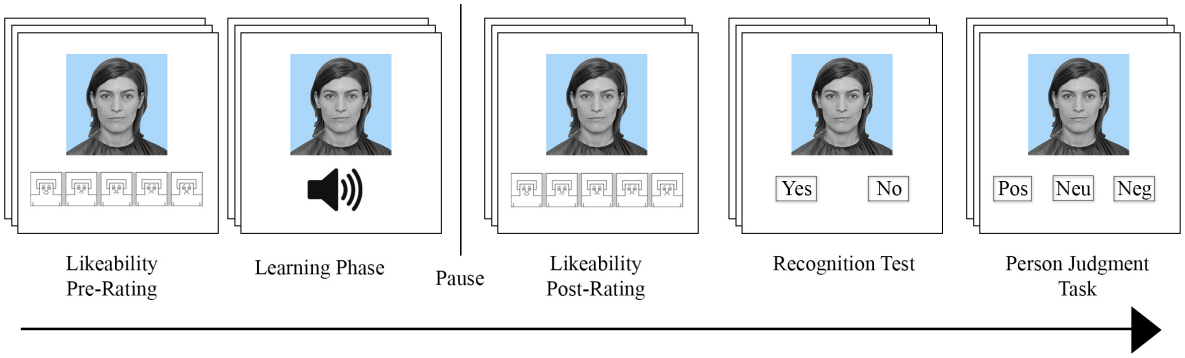

Figure 1. Schematic illustration of the experimental procedure. ERP effects were investigated during the person judgment task. *Consent to publish photo here has been given by depicted person.*

To create the global impression that all people encountered during the experiment exist in real life, we included well-known filler faces and participants were told that they will see faces, 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 likeable they spontaneously found each person on a 5-point scale (adopted from the Self-Assessment Manikin; Bradley & Lang, 1994). We call this rating "spontaneous likeability", because this rating is not based on the information, but rather on the spontaneous feeling of liking, i.e. the information was not task relevant. During learning, participants saw the face and listened to the assigned verbal information. Block wise learning familiarised 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 6 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?* (per face 4 questions in total).

The test phase always started with the likeability rating (see learning phase). A recognition test and the person judgment task followed, in a counterbalanced sequence. To test recognition, participants decided if a face was familiar from the learning phase (included 80 unfamiliar filler faces). For the person judgment, participants had to judge if 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 6 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 likeability rating, faces were presented for 1 s, followed by the rating scale until response.

After the experiment, a paper-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 8 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 likeability 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 et al., 2018).

Results

Behavioral results

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

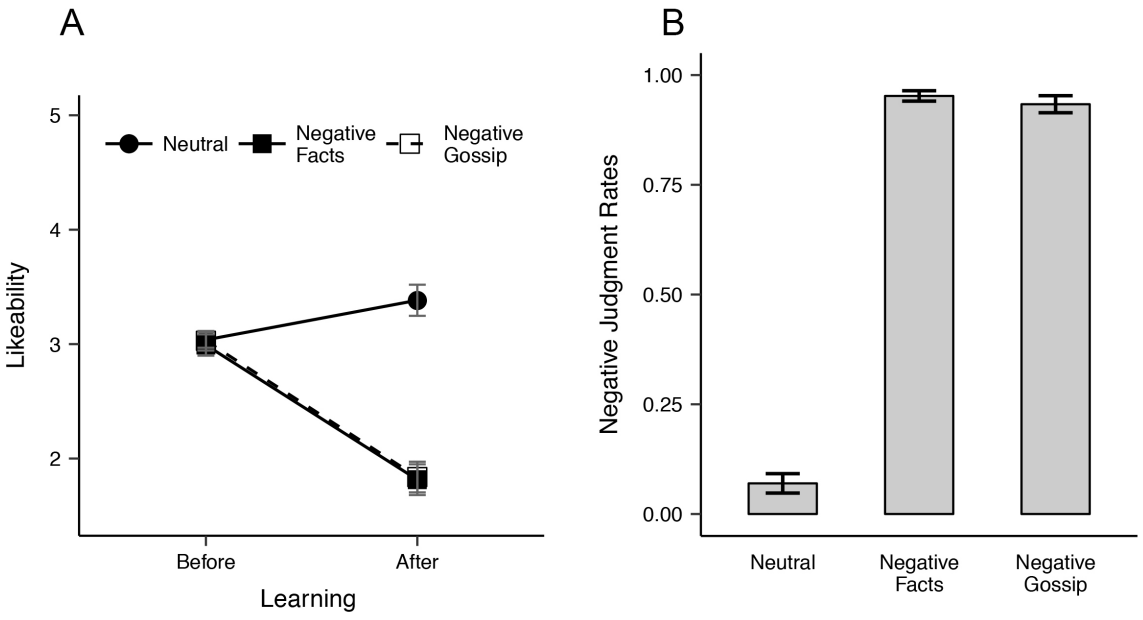

Figure 2. Behavioral results of experiment 1. Error bars show 95% confidence intervals. (A) Mean likeability rating before and after learning depending on information. (B) Mean judgment rates in the person judgment task depending on information.

Likeability rating

Whereas there was no difference in likeability 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 (2 levels: before vs. after learning), $F(1,31) = 118.57$, $p < .001$, $\eta_p^2 = .79$, $\eta_G^2 = .56$, and information (3 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 likeable 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 4.1 and 4.2).

Table 2

Means and confidence intervals for the likeability rating in experiment 1

Exp.	Phase		Neutral	Negative Facts	Negative Gossip
1	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 person judgment rates in experiment 1

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

Note. CI = confidence interval.

Table 4.1

Means and confidence intervals for reaction times during person judgment in experiment 1

Exp.	RT [ms]	Neutral	Negative Facts	Negative Gossip
1	*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 4.2

Summary of statistical results for reaction times during person judgment in experiment 1

Exp.	Source	*df*	*F*	*p*	η_G^2
1	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

ERPs

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, $F(1,31) = .60$, $p = 1$, $\eta_p^2 = .02$, $\eta_G^2 = .001$.

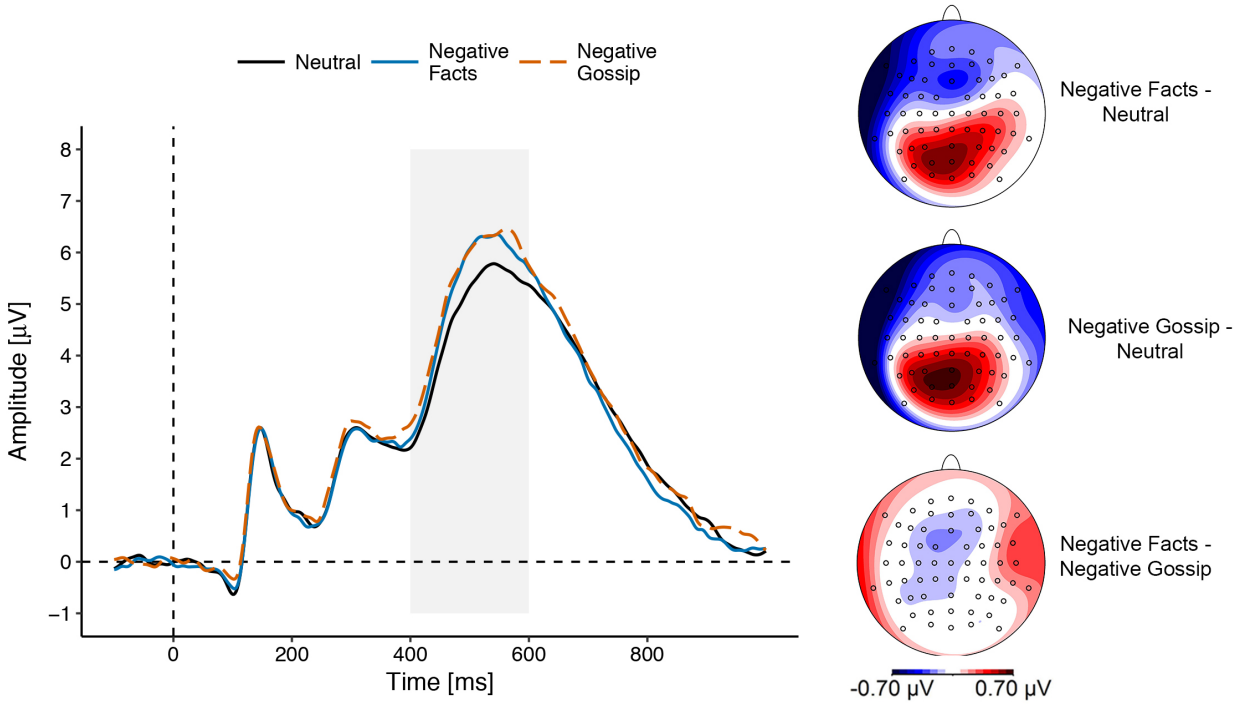

Figure 3. Grand average event-related potentials at the central-posterior site Pz show the LPP effects of information in the person judgment task in experiment 1. Scalp distributions show the effects as differences between conditions in the LPP time window 400 to 600 ms.

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 likeability, explicit person judgments and LPP amplitudes reflecting evaluative processes. Moreover, 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 likeability 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 5). 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 5

Means and confidence intervals for trustworthiness ratings in experiment 2

Exp.		Negative facts	Negative gossip	Positive facts	Positive gossip
2	*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.

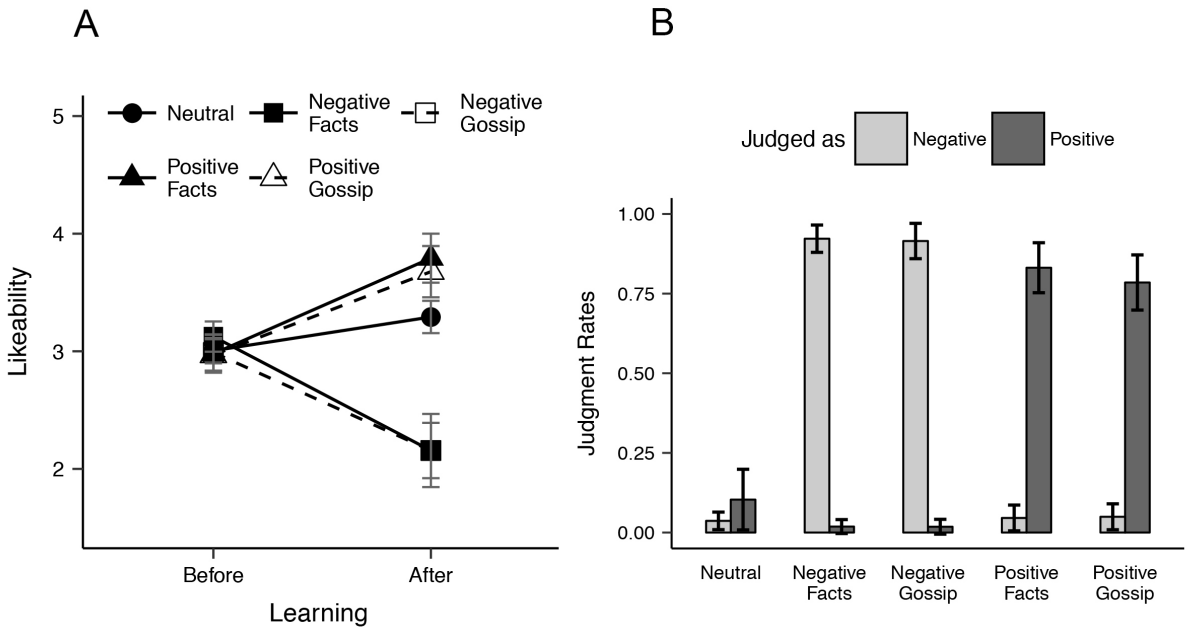

Figure 4. Behavioral results of experiment 2. Error bars are 95% confidence intervals. (A) Mean likeability ratings before and after learning depending on information. (B) Mean negative and positive judgment rates in the person judgment task depending on information.

Likeability 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 6).

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, $F(2,46) = 36.51$, $p < .001$, $\eta_p^2 = .61$, $\eta_G^2 = .31$. Likeability 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 likeable 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 likeable 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 7).

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 6

Means and confidence intervals for likeability rating for experiment 2

Exp.	Phase		Neutral	Negative Facts	Negative Gossip	Positive Facts	Positive Gossip
2							
	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 7

Means and confidence intervals for judgment rates for experiment 2

Exp. Judgment		Neutral	Negative Facts	Negative Gossip	Positive Facts	Positive Gossip
2						
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 8.1 and 8.2).

Table 8.1

Means and confidence intervals for reaction times during person judgment for experiment 2

Exp.		Neutral	Negative Facts	Negative Gossip	Positive Facts	Positive Gossip
2	*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 8.2

Summary of statistical results for reaction times during person judgment for experiment 2

Exp.	Source	*df*	*F*	*p*	η_p^2	η_G^2
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

Exp.	Source	*df*	*F*	*p*	η_p^2	η_G^2
	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

ERPs

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, $F(2,46) = 1.46$, $p = .24$, $\eta_p^2 = .060$, $\eta_G^2 = .004$, but conditions differed between 300 and 350 ms $F(2,46) = 3.77$, $p = .030$, $\eta_p^2 = .14$, $\eta_G^2 = .008$. Analyses comparing positive facts to neutral information, $F(1,23) = 5.01$, $p = .11$, $\eta_p^2 = .18$, $\eta_G^2 = .007$, and to positive gossip, $F(1,23) = 5.47$, $p = .085$, $\eta_p^2 = .19$, $\eta_G^2 = .011$, did not reach significance after Bonferroni corrections. Positive gossip did not differ from neutral information, $F(1,23) = .36$, $p = 1$, $\eta_p^2 = .015$, $\eta_G^2 < .001$.

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, $F(2,46) = 7.40$, $p = .002$, $\eta_p^2 = .24$, $\eta_G^2 = .057$. Separate analyses revealed an enhanced positivity for facts

compared to neutral information, $F(1,23) = 14.73$, $p = .003$, $\eta_p^2 = .39$, $\eta_G^2 = .074$, as well as for gossip compared to neutral information, $F(1,23) = 9.67$, $p = .015$, $\eta_p^2 = .30$, $\eta_G^2 = .061$. There was no amplitude difference between negative facts and gossip, $F(1,23) = .050$, $p = 1$, $\eta_p^2 = .002$, $\eta_G^2 < .001$.

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

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

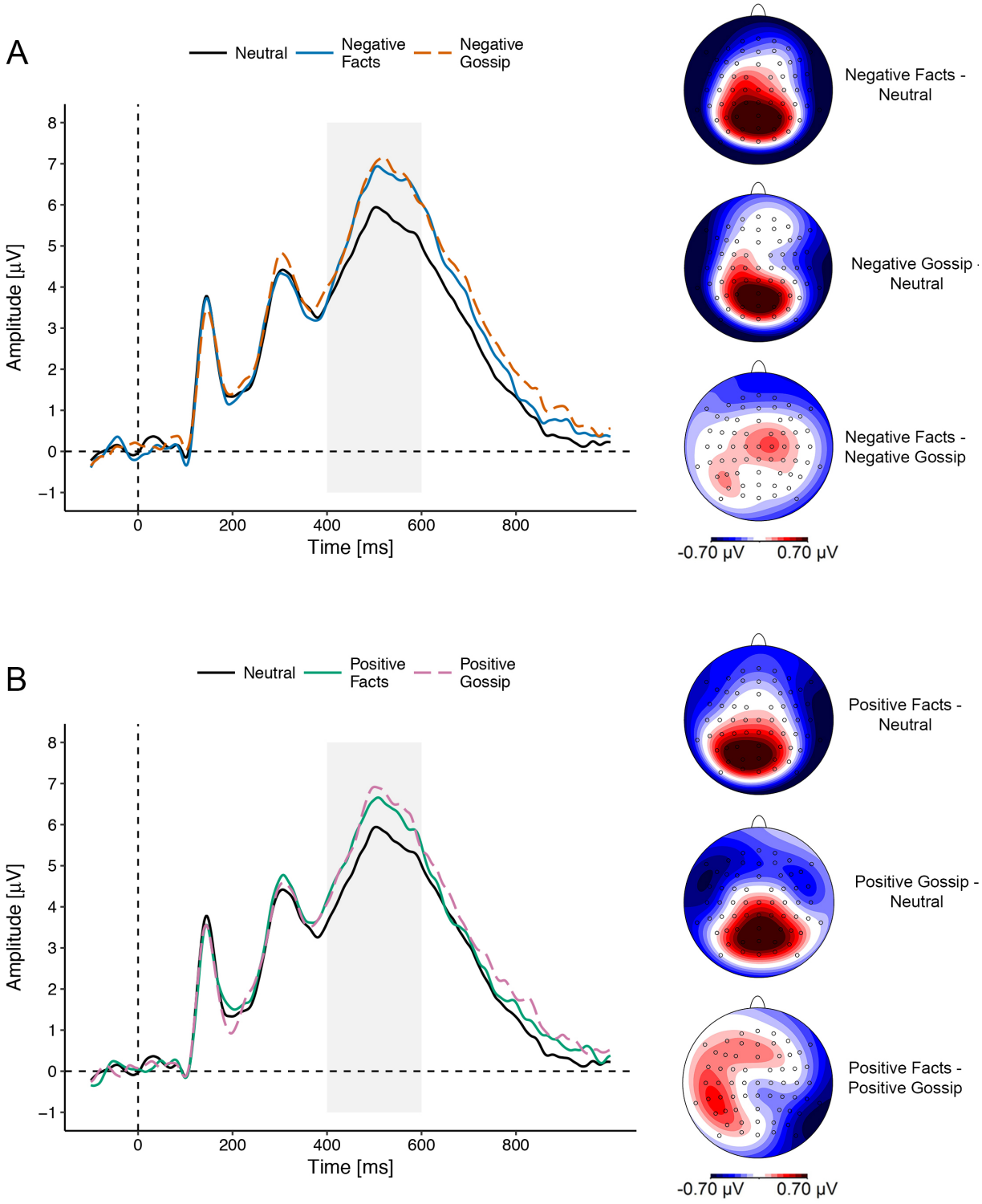

Figure 5. Grand average ERPs at the central-parietal site Pz show the LPP effects of information in the person judgment task in experiment 2. Scalp distributions show the effects as

differences between conditions in the LPP time window between 400 and 600 ms. (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 likeability, 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 we judge and emotionally evaluate others is influenced by what we know and what we 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 likeability 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”, etc. 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 et al., 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., 2017b) 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 more trustworthy than 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 more trustworthy than 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., 2017a). 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, may 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.

At 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, likeability 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, 2017; 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. (2017). Quotation and the lexicon-pragmatics interface: The case of sogenannt (‘so-called’). Submitted to *Zeitschrift für Sprachwissenschaft.* lingbuzz/003338

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. (2016). The Role of Emotion Regulation in Moral Judgment. *Neuroethics*, *1251*(1), 312–12. 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*, *42*, 11–10. 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., et al. (2017a). 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., et al. (2017b). 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

Footnote 1 (p. 5):

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.

Footnote 2 (p. 32):

“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., 2017b, p. 37)

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.