OXFORD

Attending to Eliza: rapid brain responses reflect competence attribution in virtual social feedback processing

Sebastian Schindler,[1,2,3] Gregory A. Miller,[4] and Johanna Kissler[1,2]

[1] Department of Psychology, Bielefeld University, P.O. Box 100131, 33501 Bielefeld, Germany [2]Center of Excellence Cognitive Interaction Technology (CITEC), Bielefeld University, 33619 Bielefeld, Germany [3]Institute of Medical Psychology and Systems Neuroscience, University of Münster, 48149 Münster, Germany [4]Department of Psychology and Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, P.O. Box 951563, CA 90095-1563, USA

Correspondence should be addressed to Johanna Kissler, Department of Psychology, Bielefeld University, 33501 Bielefeld, Germany.
E-mail: johanna.kissler@uni-bielefeld.de.

Abstract

In the age of virtual communication, the source of a message is often inferred rather than perceived, raising the question of how sender attributions affect content processing. We investigated this issue in an evaluative feedback scenario. Participants were told that an expert psychotherapist, a layperson or a randomly acting computer was going to give them online positive, neutral or negative personality feedback while high-density EEG was recorded. Sender attribution affected processing rapidly, even though the feedback was on average identical. Event-related potentials revealed a linear increase with attributed expertise beginning 150 ms after disclosure and most pronounced for N1, P2 and early posterior negativity components. P3 and late positive potential amplitudes were increased for both human senders and for emotionally significant (positive or negative) feedback. Strikingly, feedback from a putative expert prompted large P3 responses, even for inherently neutral content. Source analysis localized early enhancements due to attributed sender expertise in frontal and somatosensory regions and later responses in the posterior cingulate and extended visual and parietal areas, supporting involvement of mentalizing, embodied processing and socially motivated attention. These findings reveal how attributed sender expertise rapidly alters feedback processing in virtual interaction and have implications for virtual therapy and online communication.

Key words: virtual communication; EEG/ERP; social context; feedback; mentalizing networks; emotion; language

Introduction

Computer scientist Joseph Weizenbaum (1966) observed that his secretary confided in a relatively simple computer program simulating a Rogerian therapist, seemingly disregarding its artificial nature. This suggests that attributions of competence guide communication. Automated psychotherapy dates to Lang *et al.* (1970), and nowadays verbal-exchange-based psychotherapy is being successfully administered via the Internet (e.g. see Barak *et al.,* 2009). At the same time, chatbots increasingly emulate human interactive behavior (Lee, 2016) and, masked as 'humans', try to influence public opinion (Ferrara *et al.,* 2016).

These phenomena foreground the question of how, in the absence of physical cues, specific sender attributions alter the

processing of communicative content. The present research addresses this question, measuring brain responses evoked by written evaluative personality feedback given during virtual interaction by senders supposedly varying in competence. Feedback consisted of positive, neutral or negative trait adjectives. Event-related brain potentials (ERPs) elicited by these single-word messages were analyzed and their brain sources localized.

Visual ERPs elicited by single emotional words devoid of communicative context differ from those evoked by neutral words (Fischler & Bradley, 2006; Kissler *et al.,* 2007; Hofmann *et al.,* 2009; Hinojosa *et al.,* 2010). These differences are typically reflected in a larger early posterior negativity (EPN, Kissler *et al.,* 2007) and a larger late positive potential (LPP) for emotional than for neutral words (Kanske & Kotz, 2007; Herbert *et al.,* 2008; Hofmann *et al.,* 2009; Hinojosa *et al.,* 2010). While the EPN is associated with a spontaneous attention shift to emotional and arousing stimuli, likely reflecting an automatic conceptual tagging process (Schupp *et al.,* 2006; Kissler *et al.,* 2009), which has been found to be relatively robust to distracting tasks (Kissler *et al.,* 2009), the LPP is thought to be involved in explicit stimulus evaluation and affective labeling and subject to controlled attention processes (Schupp *et al.,* 2004; Hajcak *et al.,* 2009; Frenkel & Bar-Haim, 2011). On the LPP, emotional content and explicit task have sometimes (Schupp *et al.,* 2007; Schacht & Sommer, 2009a; Schindler & Kissler, 2016b), but not always (Ferrari *et al.,* 2008; Kissler *et al.,* 2009) been found to interact. Emotion effects on earlier ERPs are mixed: some studies report modulations of the P1, the N1 or the P2 component (Hofmann *et al.,* 2009; Scott *et al.,* 2009; Sass *et al.,* 2010; Kanske *et al.,* 2011; Keuper *et al.,* 2013; Keuper *et al.,* 2014), which have been related to early stages of stimulus detection and discrimination (P1, N1) and attention and lexical access (P2).

However, various communication models agree that semantic significance is an emergent property of the entire communicative situation (Blumer, 1969) and that sender attributes are critical in constructing such significance (Shrauger & Schoeneman, 1979). This view resonates with a more recent surge of interest in predictive mechanisms in psychology and neuroscience (e.g. Bar, 2007; Gilbert & Wilson, 2007). Accordingly, studies have begun to address context effects in emotion processing (Barrett *et al.,* 2007; Barrett *et al.,* 2011). Emotional or self-referential contexts established by short sentences have been found to modulate face processing (Diéguez-Risco *et al.,* 2013; Wieser *et al.,* 2014; Diéguez-Risco *et al.,* 2015; Klein *et al.,* 2015; Aguado *et al.,* 2019 ; Li *et al.,* 2019). Typically, congruence effects occur on the LPP, for example showing that negative context amplifies LPP amplitudes in response to negative expressions (Diéguez-Risco *et al.,* 2013; Diéguez-Risco *et al.,* 2015; Aguado *et al.,* 2019). However, earlier modulations can also be observed, specifically for self-referential context, increasing EPN responses towards emotional expressions (Aguado *et al.,* 2019; Li *et al.,* 2019). In the area of emotional language, context variations include a speaker's gaze direction (Rohr & Abdel Rahman, 2015) or self-relevance induced via pronouns (Herbert *et al.,* 2011) or sentence addressee (Fields & Kuperberg, 2012; Bayer *et al.,* 2017), as well as inferences about socio-emotional responses from prototypical vignette scenarios (Leuthold *et al.,* 2012).

Targeting the issue of perceived sender identity in virtual interaction as one important social context variable, in a recent scenario we told participants that they were interacting with either a human or a computer from whom they would receive written personality feedback. The computer was introduced as randomly acting (Schindler *et al.,* 2015; Schindler & Kissler, 2018) or as socially intelligent (Schindler & Kissler, 2016a). Both manip-

ulations resulted in higher ERP amplitudes for feedback from supposed human senders, appearing as early as the P2 component, although participants received, in fact, random and identical feedback in both conditions. Positive and negative feedback also elicited larger brain responses than did neutral feedback. Moreover, processing of emotionally significant feedback was boosted when it was perceived as human-generated, as evident from interactions on the EPN (Schindler *et al.,* 2015; Schindler & Kissler, 2016a) and P3 (Schindler & Kissler, 2016a; Schindler & Kissler, 2018). However, human–computer differences decreased when the computer was said to be socially intelligent, indicating that factors other than human–computer differences affect content processing and help set processing priors in social interaction. Whether ERP responses in virtual interaction contexts differ depending on social attributes of putative human senders has not yet been investigated.

In a social psychology experiment, Collins & Stukas (2006) had participants complete a short version of the Eysenck Personality Inventory, after which they received personality feedback from either high- or low-status therapists. Feedback that was inconsistent with the participants´ self-concept was more likely accepted when coming from high-status therapists. Thus, anticipated sender expertise leads recipients to ascribe more validity to the perceived feedback.

Indeed, updating and integration of feedback from others with one's self-concept are key aspects of communication (Shrauger & Schoeneman, 1979), likely activating mentalizing, self-reflection and embodied processing. Social neuroscience studies have identified the dorsomedial prefrontal cortex as a key mentalizing region that interacts with the posterior cingulate cortex (PCC), which is thought to mediate self-reflective processes (Lieberman, 2007). Mentalizing and embodied word processing also activate the somatosensory cortices (e.g. Jacoby *et al.,* 2016; Pulvermüller, 2005), whose responses are further modulated by affective significance (Gazzola *et al.,* 2012). Moreover, when processing social feedback, superior medial frontal activations are increased as a correlate of mentalizing about human interaction partners (Hughes & Beer, 2012; Korn *et al.,* 2012). Accordingly, source reconstructions of activation differences between human and computer senders revealed superior frontal and somatosensory cortex effects (Schindler & Kissler, 2016a) as well as large visual cortex activations. A recent fMRI-adapted variant of our original paradigm also revealed activations in superior medial prefrontal cortex, PCC and somatosensory structures (Schindler et al., 2019). The visual cortex activations have been suggested to index higher attention to messages from the human sender (Schindler *et al.,* 2015), in line with the motivated attention model (Lang *et al.,* 1997), extending it from emotional content to contextual social significance.

The combined findings from social psychology and neuroscience imply that larger ERPs should be elicited by feedback from senders perceived as more competent, reflecting increased activity of structures involved in attention, mentalizing and updating of one's self-concept. To address this, we investigated whether and when sender expertise amplifies processing of language-based feedback in a virtual interaction scenario. In different experimental blocks, participants were told that they would receive personality feedback by an expert (a psychotherapist), a layperson or a randomly acting computer. Based on previous research, we expected higher ERP amplitudes for allegedly human than for allegedly computer feedback. We hypothesized this difference to be larger when supposedly interacting with an expert and investigated the timing of any effects. We also expected emotional content to impact feedback processing. In

Table 1. Comparisons of positive, neutral and negative adjectives by one-way ANOVAs

Variable	Positive adjectives (n = 70)	Neutral adjectives (n = 40)	Negative adjectives (n = 70)	F (2,147)
Valence	7.34[a] (0.63)	4.94[b] (0.28)	2.85[c] (0.67)	1016.25***
Arousal	4.66[a] (0.76)	3.2[b] (0.82)	4.78[a] (0.74)	60.96***
Concreteness	2.86[a] (1.01)	5.11[b] (1.51)	3.18[a] (0.66)	65.70***
Word length	9.30 (2.94)	8.95 (2.43)	8.79 (2.65)	0.64
Word frequency (per million)	493.69 (780.45)	512.60 (703.15)	483.43 (769.05)	0.02
Familiarity (absolute)	39934.16 (17585.69)	23488.33 (10506.85)	30036.70 (14497.37)	0.59
Regularity (absolute)	265.70 (423.44)	103.85 (186.28)	208.61 (406.98)	2.35
Neighbors Coltheart (absolute)	4.60 (6.54)	2.38 (2.95)	3.21 (3.85)	2.88
Neighbors Levenshtein (absolute)	7.47 (8.31)	4.70 (3.73)	5.86 (6.06)	2.38

Note: ***$P \leq 0.001$. Standard deviations appear in parentheses below means; means in the same row sharing the same superscript letter do not differ significantly from one another at $P \leq 0.05$; means that do not share subscripts differ at $P \leq 0.05$ based on LSD test post hoc comparisons

single-word processing studies (e.g. Kissler *et al.,* 2007; Kissler *et al.,* 2009; Schacht & Sommer, 2009a; Schacht & Sommer, 2009b), emotion effects have been most clearly seen on the EPN and to some extent also on the LPP component. However, previous studies with the present feedback design have revealed more variability with regard to their timing (cf. Schindler *et al.,* 2015; Schindler & Kissler, 2016a, 2018), leading us to expect higher amplitudes for emotionally valent feedback without clear predictions about the timing of their onset. Overall, we analyzed the sequence of typical visual ERP components (N1, P2, EPN, P3 and LPP) with analyses of variance and evaluated the direction of any differences with orthogonal polynomial (linear, quadratic) trend analysis. Significant effects were further analyzed in source space to identify their cerebral generators.

Materials and Methods

Participants

To obtain a sample size comparable to previous feedback studies (Schindler & Kissler, 2016a; Schindler & Kissler, 2018), 39 undergraduates from Bielefeld University participated after providing written informed consent in a study approved by the local Ethics Committee. They received 14 Euros for participation. Seven participants were excluded due to large artifacts or technical problems including a fire alarm. One participant was excluded due to a reported acute anxiety disorder, and one because of confusion about the condition-run assignment. The resulting 30 participants (23 females) were 22.03 years of age on average (s.d. = 3.73), right-handed and had normal or corrected-to-normal vision. For these, screenings with the Beck Depression Inventory and the State and Trait Anxiety Questionnaire (BDI, STAI: Beck *et al.,* 2001; Spielberger *et al.,* 1999) revealed no clinically relevant depression (M = 4.87, s.d. = 5.02) or anxiety scores (M = 37.60, s.d. = 9.54).

Stimuli

The stimuli of Schindler & Kissler (2018) were used. These had been rated by 22 students, who did not participate in the ERP experiment, in terms of valence, arousal and concreteness. Raters had been instructed to consider the adjectives' valence and arousal in an interpersonal evaluative context. The selected 180 adjectives (70 positive, 40 neutral, 70 negative) were matched in their linguistic properties, such as word length, frequency, familiarity and regularity (see Table 1). Positive and negative adjectives differed only in valence, while neutral adjectives were allowed to deviate on rated concreteness (see Table 2), since truly neutral evaluative trait adjectives are rare.

Procedure

Participants were told that after a self-introduction, they would be evaluated by two unknown raters, one putatively an expert psychotherapist and one a layperson, and a randomly operating computer algorithm (within-design). The sequence of 'human sender' conditions was counterbalanced across participants, while the computer feedback was always in between, supposedly enabling the experimenter to switch judges in the adjacent laboratory room.

Upon arrival, participants were instructed to briefly describe themselves in a structured interview allegedly videotaped for both human judges to convey an impression of the participant, along with a short personality inventory that participants filled out. They also completed a demographic questionnaire, BDI and STAI (see above). As a manipulation check, participants also rated general expertise and competence of psychotherapists' and laypeople to evaluate others (see below and the Supplementary data). A research assistant left the testing room 15 min ahead of the fictitious feedback, guiding an 'unknown person' to a laboratory room next to the testing room to foster face validity.

Stimuli were presented by software described as 'Interactional Behavioral Systems' (implemented in Presentation software; www.neurobehavioralsystems.com) supposedly allowing instant online communication. The feedback was randomly generated in all conditions, but the 10 most positively rated and 10 most negatively rated adjectives were always rejected to increase face validity (e.g. brutal, visionary). Overall, 40 affirmative positive, 40 neutral and 40 affirmative negative feedback trials were administered. Color change (blue or purple, counterbalanced) of the presented adjective indicated whether or not the respective adjective applied to the participant. In the human condition, a color change between 1500 and 2500 ms after adjective onset indicated a decision by the respective interaction partner. This simulated varying feedback latencies in humans. In the computer condition, color changes occurred between 1400 and 1600 ms. Color changes always lasted for 1000 ms, followed by a fixation cross for 1000 to 1500 ms. Figure 1 illustrates the procedure. Participants completed a post-experimental questionnaire about the experiment in general and feedback appropriateness. One of the final 30 participants reported spontaneously not to believe the presence of other participants (see Supplementary data). These data were nevertheless retained.

G recording and analysesEEG was recorded from 128 BioSemi active electrodes (www.biosemi.com) at 2048 Hz. During recording, the common mode sense active electrode and the driven right leg passive electrode were used as reference and
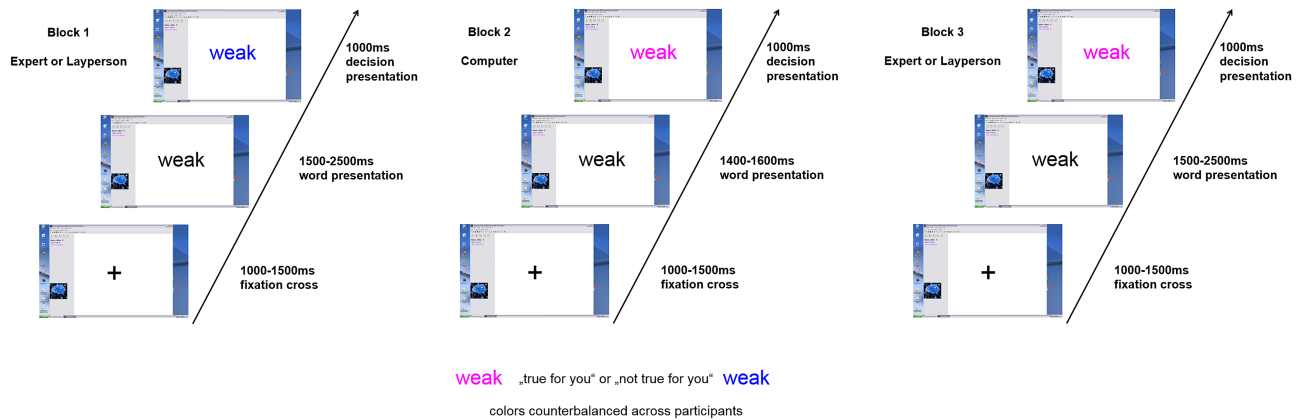
Fig. 1. Outline of the experimental design: Three different blocks of single word feedback presentation were used. Block 1 and 3 were counterbalanced and represented "expert" or "layperson" feedback, respectively, whereas Block 2 always represented computer feedback. Color change represents feedback applicability Color-significance assignment was counterbalanced across participants.

ground electrodes, respectively (www.biosemi.com/faq/cms&drl.htm). Horizontal and vertical electrooculograms (EOGs) were monitored using four facial electrodes placed near the outer canthi and on the cheek below each pupil and used for EOG correction and artifact rejection.

Pre-processing was performed using SPM8 for EEG data (http://www.fil.ion.ucl.ac.uk/spm/), and statistical analyses of the ERP data were performed using EMEGS (Peyk *et al.,* 2011). Data were down-sampled to 250 Hz and band-pass-filtered from 0.16 to 30 Hz with a fifth-order Butter worth zero-phase-shift filter as applied in previous feedback studies (e.g. Schindler *et al.,* 2015; Schindler & Kissler, 2016a; Schindler & Kissler, 2018) to avoid spillover of effects from the word pre-viewing phase to the feedback phase. Filtered data were segmented from 500 ms before stimulus onset until 1000 ms after stimulus presentation. Results are presented without baseline correction in order to avoid introduction of pre-baseline differences into the feedback phase. However, there were no apparent differences in the 200 ms immediately preceding the color change (see Figures 2 and 3), and control analyses with baseline correction lead to analogous results. For trials with individual channels exceeding a threshold of 160 µV peak-to-peak, automatic artifact detection was used. Data were averaged, using a robust averaging algorithm, excluding possible further artifacts (Litvak *et al.,* 2011). Overall, 6.61% of all electrodes were interpolated and 18% of all trials were rejected, leaving on average 32.81 trials per condition. There were no differences in the number of rejected trials as a function of sender ($F_{(2,58)} = 1.30, P = 0.28$, partial $\eta^2 = 0.04$), feedback content ($F_{(2,58)} = 1.75, P = 0.18$, partial $\eta^2 = 0.06$) or their interaction ($F_{(4,116)} = 0.58, P = 0.68$, partial $\eta^2 = 0.02$).

ERP components were scored in time windows from 180 to 220 ms to investigate N1 effects, 150 to 200 ms for P2, 250 to 350 ms for EPN, 300 to 400 ms for P3, 400 to 650 ms for early LPP and 650 to 900 ms for late LPP effects (see also Schindler & Kissler, 2016a). For the N1 and EPN time windows, two symmetrical occipital clusters of nine electrodes each were examined (left: I1, OI1, O1, PO9, PO9h, PO7, P9, P9h, P7; right: I2, OI2, PO10, PO10h, PO8, P10, P10h, P8). For the P2, P3 and LPP time windows, a large central cluster was investigated (26 electrodes: FCC1h, FCC2h, C3h, C1, C1h, Cz, C2h, C2, C4h, CCP3h, CCP1, CCP1h, CCPz, CCP2h, CCP2, CCP4h, CPz, CPP1, CPPz, CPP2, P1h, Pz, P2h, PPO1h, PPOz, PPO2h; e.g. Schindler *et al.,* 2015, 2016a, 2018).

Source reconstructions for time windows with significant ERP differences were computed and statistically assessed with SPM8 for EEG (Litvak & Friston, 2008). A realistic boundary element head model (BEM) was derived from SPM's template head model based on a Montreal Neurological Institute (MNI) brain. Electrode positions were transformed to match the template head, which generates reasonable results even when an individual subject's head differs from the template (Litvak *et al.,* 2011). Average electrode positions as provided by BioSemi were co-registered with the cortical mesh template which was used to calculate the forward solution. The inverse solution was calculated via group inversion (Litvak & Friston, 2008) and SPM's multiple sparse priors algorithm applied.

Statistical analyses

Three (sender: human expert, computer, layperson) by three (emotion: negative, neutral, positive) repeated-measures ANOVAs were calculated for time windows and electrode clusters of interest. To follow up on significant effects, we tested the hypothesis that feedback from the expert should induce the largest amplitude increase, followed by the layperson, and finally the computer. Therefore, linear (expert 1, layperson 0 and computer −1) and quadratic (expert 1, layperson −2 and computer 1) trends were calculated. For emotion effects, amplitude increase in response to both positive and negative feedback was anticipated. To test this, again, linear (positive 1, neutral 0, negative −1) and quadratic trends (positive 1, neutral −2, negative 1) were compared to identify the shape of the effects. Partial eta-squared (partial η^2) was estimated to describe effect sizes (Cohen, 1988). When Mauchly's test indicated a violation of sphericity, degrees of freedom were corrected according to Greenhouse–Geisser, reporting corrected P values and effect sizes.

Statistical analyses in source space were restricted to time windows that were significant at the scalp and focused on the same contrasts. 3D reconstructions were generated as NIFTI images (voxel size = 2 mm ∗ 2 mm ∗ 2 mm), which were smoothed using an 8 mm full-width half-maximum filter. Similar to previous studies (Schindler & Kissler, 2016b), we describe statistical differences in source activity of voxels differing at least at an uncorrected threshold of $P < 0.001$ and a minimum of 25 significant voxels per cluster. Family-wise error (FWE)-corrected results with a threshold of $P < .05$ are also reported in all tables. Activated brain regions were identified using the LONI atlas (Shattuck *et al.,* 2008).
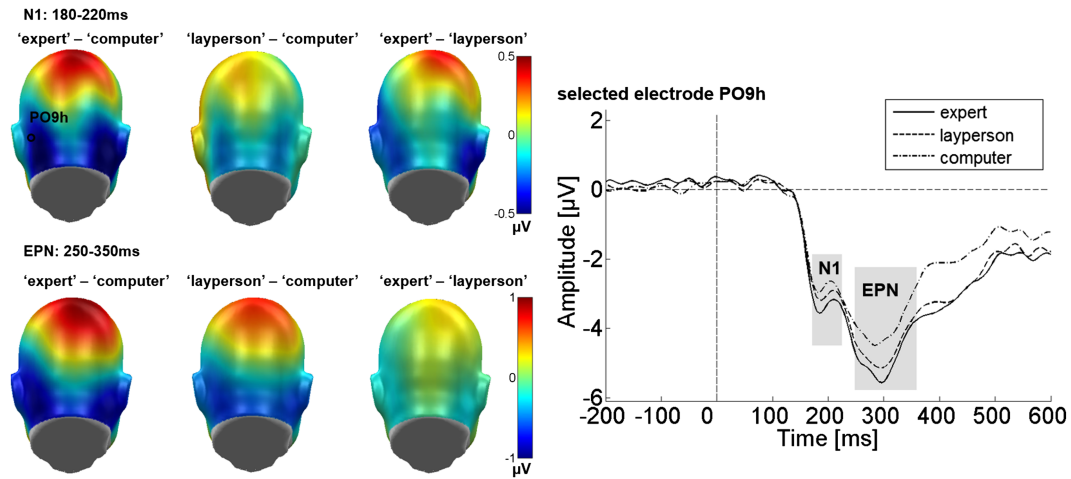
Fig. 2. Difference topographies for the occipital electrode cluster in the N1 and EPN time windows. Blue color indicates more negativity and red color more positivity for the respective comparison. Waveforms for electrode PO9h illustrate the ERP time courses for the three senders over left-occipital areas.
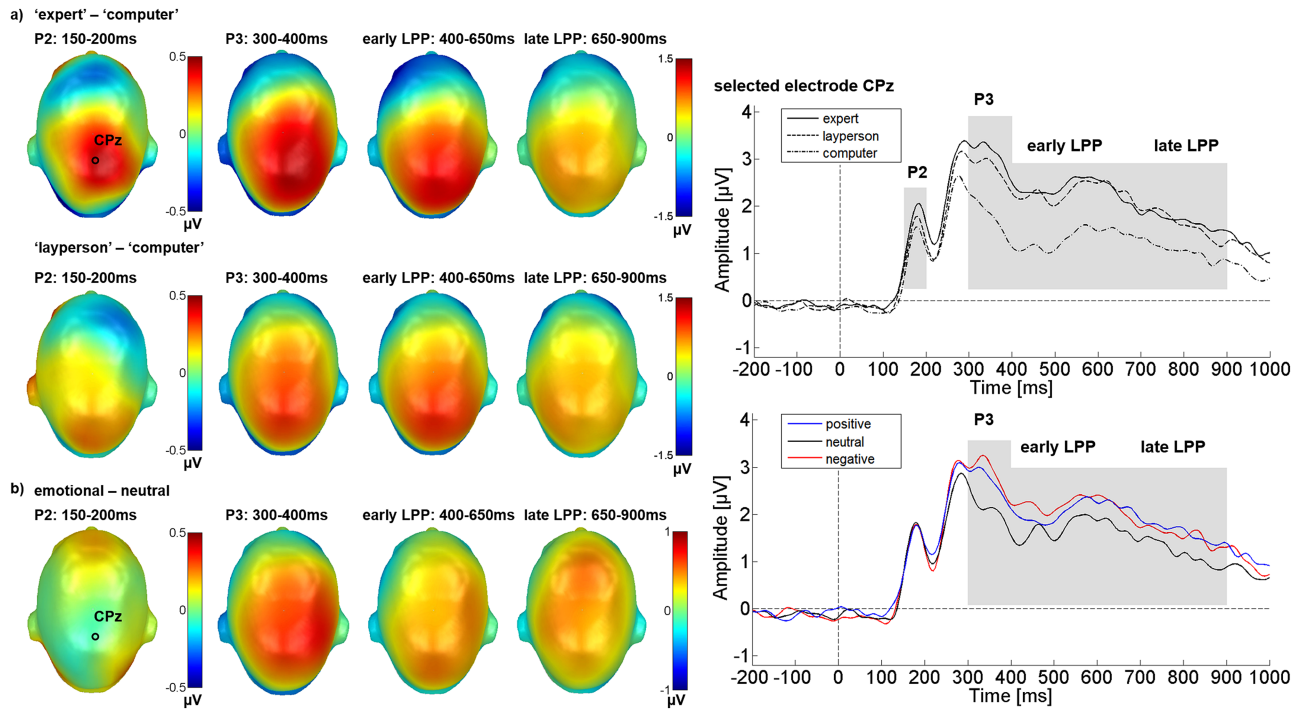

Fig. 3. Centro-parietal effects of sender and emotion. (A) Sender difference topographies: red color indicates more positivity, and blue color indicates more negativity for the respective human sender. (B) Emotion difference topographies: red indicates more positivity, and blue indicates more negativity for emotional feedback. ERP waveforms for electrode CPz show the time courses for sender and emotion main effects.

Results

Manipulation check

Before testing, participants answered four questions regarding their beliefs about psychotherapists' ability to evaluate others (for details, see Supplementary data). Participants indeed rated psychotherapists as more competent than laypersons in their ability to evaluate others ($M_{diff} = 3.08$, s.d. = 0.50). Post-experimental questioning about the overall appropriateness of feedback decisions showed a main effect of sender ($F_{(2,56)} = 9.19$, $P < 0.001$, partial $\eta^2 = 0.247$). Although, in fact on average identical, expert ($M = 3.24$, s.d. = 0.74) and layperson feedback ($M = 3.03$, s.d. = 0.68) was rated more appropriate than computer feedback ($M = 2.52$, s.d. = 0.69; $P = 0.001$ and 0.007). The linear

contrast explained rated sender feedback appropriateness best ($F_{(1, 28)} = 13.39$, $P = 0.001$; 94% of the sender variance), while the quadratic contrast was not significant ($F_{(1, 28)} = 1.50$, $P = 0.231$), the latter indicating that laypersons were judged intermediate in expertise.

N1

As illustrated in Figure 2, the occipital sensor cluster showed a main effect of sender ($F_{(2,58)} = 3.27$, $P = 0.045$, partial $\eta^2 = 0.101$). The linear contrast revealed that feedback by the expert elicited more negativity than did feedback by the computer ($F_{(1, 29)} = 6.00$, $P = 0.021$; 99% of the sender variance). The absence of a quadratic effect ($F_{(1, 29)} = 0.047$, $P = 0.830$) confirmed a linear increase with
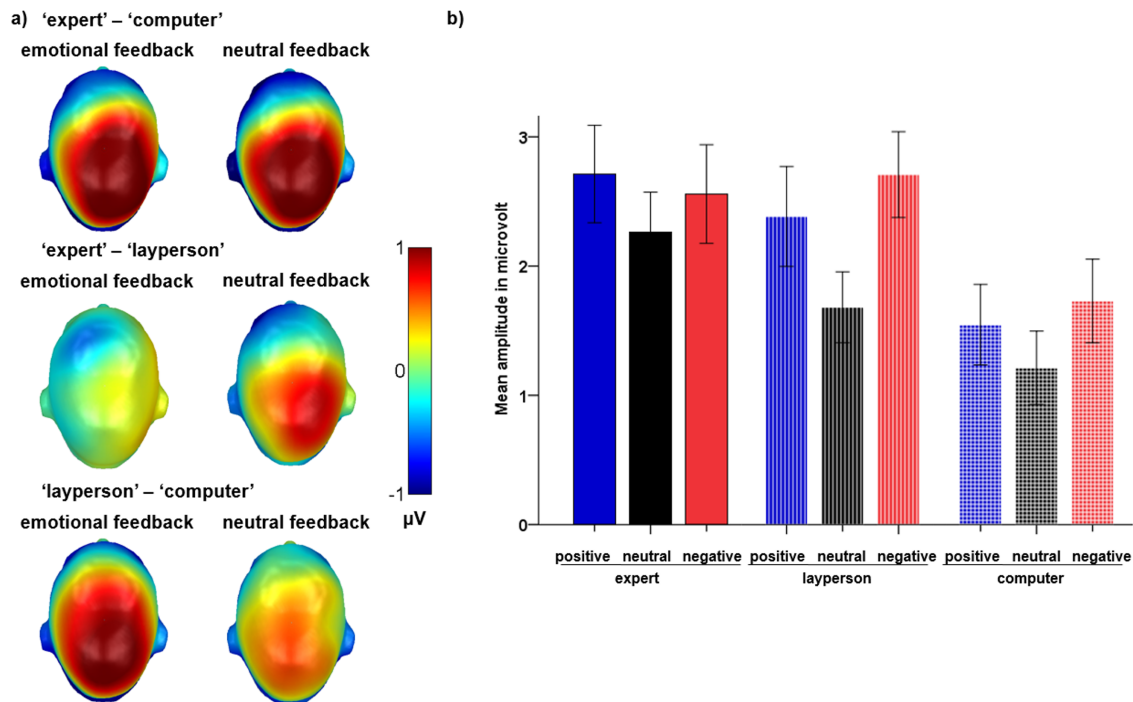
Fig. 4. Interaction between sender and emotion in the P3 time window. (A) Difference topographies comparing amplitudes elicited by neutral feedback (left panel) and emotional feedback (right panel, across negative and positive) for the indicated sender pairs. Red color indicates more positivity and blue color more negativity. (B) Mean amplitudes in microvolts over the centro-parietal sensor cluster are displayed for all types of feedback. Error bars are ± 1 SEM. Whereas the two human senders did not differ from each other in the response to emotional feedback they elicited, the expert elicited stronger responses to neutral feedback.

increasing attributed expertise (see right panel of Figure 2). No other effects approached significance.

P2

For P2, the central region illustrated in Figure 3 showed a main effect of sender ($F_{(2,58)} = 3.23$, $P = 0.047$, partial $\eta^2 = 0.100$). Again, the linear contrast revealed that feedback from the expert elicited more positivity than did feedback from the computer ($F_{(1,29)} = 6.00$, $P = 0.021$; 99% of the sender variance). The quadratic contrast was not significant ($F_{(1,29)} = 0.051$, $P = .823$), supporting a linear amplitude increase with attributed expertise. The main effect of emotion and its interaction with sender did not approach significance.

EPN

Consistent with Figure 2, a sender effect was found from 250 to 350 ms ($F_{(2,58)} = 7.62$, $P = 0.001$, partial $\eta^2 = 0.208$). The linear contrast revealed that feedback from the expert elicited more negativity than did computer feedback ($F_{(1,29)} = 9.43$, $P = 0.005$; 88% of the sender variance). The absence of a clear quadratic effect ($F_{(1,29)} = 3.18$, $P = 0.085$; 12% of the sender variance) confirms a linear increase with increasing attributed expertise. No other effects were clearly significant (Fs < 2.72, P > 0.074).

P3

A central region was responsive between 300 and 400 ms to both sender ($F_{(2,58)} = 12.867$, $P < 0.001$, partial $\eta^2 = 0.307$) and emotional content ($F_{(2,58)} = 12.874$, $P < 0.001$, partial $\eta^2 = 0.307$; see Figure 3). Feedback by the expert elicited more positivity than that by the computer sender ($F_{(1,29)} = 15.64$, $P < 0.001$; linear trend 92% of the sender variance). The marginal quadratic contrast ($F_{(1,29)} = 4.13$, $P = 0.051$; 8%) indicates that the layperson fell between the other

two but closer to the expert (see Figure 3, upper right). For the emotion main effect ($F_{(2,58)} = 12.874$, $P < 0.001$, partial $\eta^2 = 0.307$), the quadratic contrast confirmed the impression in the lower-right panel of Figure 3 and the right panel of Figure 4 that both positive and negative feedback elicited larger P3 responses than did neutral feedback ($F_{(1,29)} = 25.98$, $P < 0.001$; 97% of the emotion variance). The linear contrast not approaching significance ($F_{(1,29)} = 0.81$, $P = 0.376$) indicates that positive and negative did not differ.

An interaction between sender and emotion ($F_{(4,116)} = 2.51$, $P = 0.046$, partial $\eta^2 = 0.080$; see Figure 4) indicated that processing of feedback content differed depending on its source. Further trend analysis across all three senders revealed a significant quadratic by quadratic interaction ($F_{(1,29)} = 5.04$, $P = 0.033$; 61% of the sender by emotion interaction), suggesting that the above reported quadratic emotion effect differed between the senders. In effect, whereas processing of emotional feedback increased in a quadratic manner for both human senders, neutral feedback increased linearly ($F_{(1,29)} = 17.92$, $P < 0.001$; 99% of the variance of the sender effect on neutral feedback), rather than in a u-shape (quadratic P > 0.75), revealing selective processing of neutral content when coming from the expert (see also Figure 4).

LPP

A central sensor cluster was responsive to sender information and emotional content, resulting in main effects of sender ($F_{(2,58)} = 12.60$, $P < 0.001$, partial $\eta^2 = 0.302$) and emotion ($F_{(2,58)} = 10.81$, $P < 0.001$, partial $\eta^2 = 0.272$; see Figure 3) in the early LPP (400–650 ms). The linear contrast revealed that feedback by the expert elicited a stronger positivity than did feedback by the computer ($F_{(1,29)} = 15.40$, $P < 0.001$; 83% of the sender variance), but the quadratic contrast was also significant ($F_{(1,29)} = 6.57$, $P = 0.016$; 17%), suggesting no substantial difference between
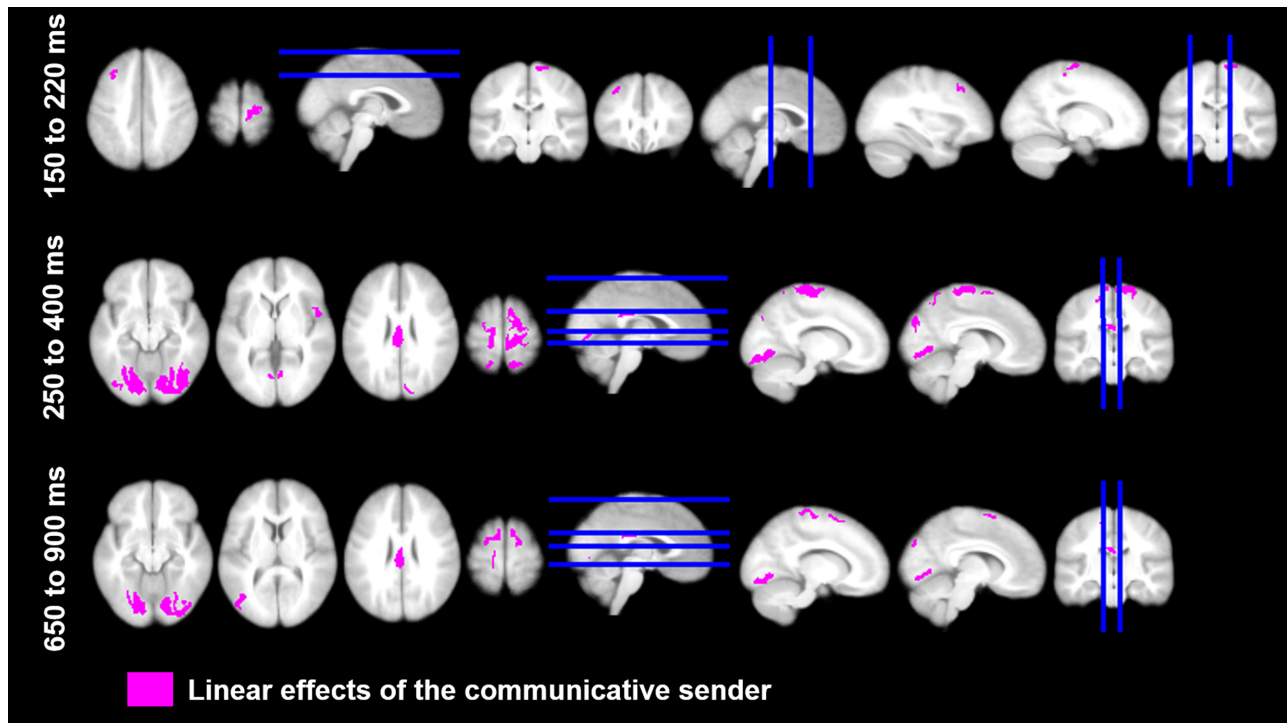
Fig. 5. Source estimations for effects of the human expertise (post hoc *t*-tests are displayed, $P < .001$).

the two human senders. Emotion ($F_{(2,58)} = 10.81$, $P < 0.001$, partial $\eta^2 = 0.272$) provided only a quadratic contrast, with positive and negative feedback eliciting larger responses than neutral feedback ($F_{(1, 29)} = 27.43$, $P < 0.001$; 98%; linear contrast $F_{(1, 29)} = 0.32$, $P = 0.572$). There was no interaction between sender and emotion.

As shown in Figures 3 and 4, the amplitude of the late LPP (650–900 ms) also responded to sender identity and emotional content of the feedback, as reflected in main effects of sender ($F_{(2,58)} = 7.27$, $P = 0.002$, partial $\eta^2 = 0.200$) and emotional content ($F_{(2,58)} = 14.77$, $P < 0.001$, partial $\eta^2 = 0.337$). The linear contrast revealed that feedback by the expert elicited a stronger positivity than did the computer sender ($F_{(1, 29)} = 10.13$, $P = 0.003$; 75% of the sender variance), whereas the quadratic contrast was marginal ($F_{(1, 29)} = 3.93$, $P = 0.057$; 25%), suggesting that the layperson fell between the other two but closer to the expert. Emotion provided a quadratic contrast ($F_{(1, 29)} = 26.45$, $P < 0.001$, 96% of the emotion variance; linear $F_{(1, 29)} = 1.25$, $P = 0.27$). There was no interaction between sender and emotion.

Source reconstruction

The significant findings on the scalp guided source-space analyses, leading us to focus on linear effects of sender in the N1/P2, EPN/P3 and late LPP time windows, quadratic effects of emotion in the P3 and LPP time windows and the sender-dependent difference for neutral feedback in the P3 time window. Since source estimations use all sensors per time window, partly overlapping time windows were collapsed to avoid multiple statistical tests of overlapping data and to obtain more concise results.

Accounting for linear effects of sender in the N1/P2 time window, the expert was found to elicit more activity in precentral and left middle frontal areas (see Figure 5, Table 2). Later in the EPN/P3 and late LPP time windows, linear sender effects revealed more activity in broad visual, parietal, frontal and somatosensory regions as well as the posterior cingulate (see Figure 5, Table 2). In particular, the expert elicited large and sustained activity in broad superior frontal regions. Control analyses (not shown) revealed that in the N1/P2 window the layperson–computer contrast was not associated with any specific supra-threshold activity in source space, whereas in later time windows the layperson–computer contrast fell within the same areas as did the expert–computer contrast, albeit less pronounced, in line with the notion of a linear activity increase with expertise in these areas. The quadratic contrast of expert and layperson *vs* computer did not reveal any supra-threshold activity.

For the emotion main effect, significant differences were found in the P3 and LPP time windows (see Figure 6 and Table 3). According to quadratic contrasts analogous to those in scalp space, visual cortex as well as parietal, posterior cingulate and left superior frontal areas were more active for emotional feedback.

Source estimations of the P3 scalp–interaction showed that neutral feedback by the expert induced more activity in the left medial superior frontal gyrus than did neutral feedback by the layperson and by the computer. No differences were observed between neutral feedback from the layperson and the computer or between emotional feedback from the senders (see Figure 7, Table 4).

Discussion

The present study investigated the effect of attributed sender expertise on processing of socio-emotional verbal feedback in a virtual communication scenario. Participants expected feedback from a psychotherapist (expert), an unspecified other person (layperson) or a randomly acting computer. In fact, all conditions were computer-generated and on average identical. This allowed

Table 2. Source estimations for linear sender effects

Cluster-level	Peak-level		MNI coordinates			LONI
Number of significant voxels	Peak t (1, 261)	Peak P-unc	x (mm)	y (mm)	z (mm)	area
N1/P2 time window (150–220 ms)						
Expert > computer						
66	3.33	<0.001	−30	24	48	Mid frontal G L
155	3.40	<0.001	22	−14	68	Precentral G R
EPN/P3 time window (250–400 ms)						
Expert > computer						
1303 (561[a])	6.03	<0.001	10	−72	−4	Lingual G R
978 (506[a])	5.89	<0.001	−20	−72	−12	Inf occipital G L
204 (147[a])	5.42	<0.001	0	−32	24	Cingulate G R
965 (325[a])	4.97	<0.001	22	−16	72	Precentral G R
545	4.36	<0.001	−20	−14	70	Precentral G L
308	4.32	<0.001	18	−82	42	Sup occipital G R
179	3.87	<0.001	16	−56	68	Sup parietal G R
47	3.69	<0.001	−6	−86	34	Cuneus L
57	3.52	<0.001	52	4	2	Precentral G R
36	3.48	<0.001	−18	−54	68	Sup parietal G L
78	3.44	<0.001	−38	−70	12	Mid occipital G L
Late LPP time window (650–900 ms)						
Expert > computer						
785 (252[a])	5.19	<0.001	12	−80	−12	Lingual G R
431 (139[a])	5.06	<0.001	−18	−76	−14	Inf occipital G L
200	3.93	<0.001	20	−82	40	Mid occipital G R
146	3.80	<0.001	0	−32	24	Cingulate G L
156	3.61	<0.001	−46	−80	12	Mid occipital G L
155	3.61	<0.001	−18	8	62	Sup frontal G L
124	3.49	<0.001	12	6	66	Sup frontal G R
76	3.30	<0.001	−14	−24	66	Precentral G L

Notes. [a]Resulting cluster size when the FWE-corrected threshold of $P < .05$ (≥25 significant voxels) was used. Number of significant voxels = the number of voxels that differ significantly between both conditions. Peak P-unc = uncorrected P value. For each significant peak, respective coordinates (x, y and z) are displayed in MNI space. A cluster may exhibit more than one peak, while only the largest peak is reported. Area = peak-level brain region as identified by the LONI atlas. R/L = laterality right or left. G = gyrus; Inf = inferior, Mid = middle, Sup = superior
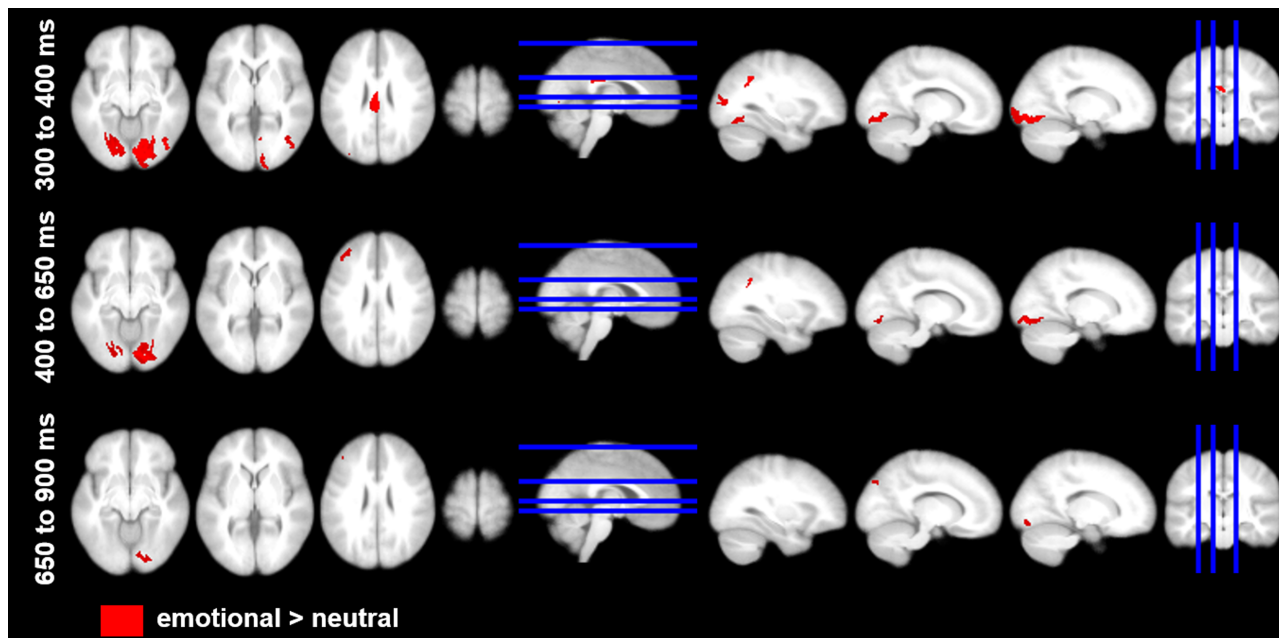
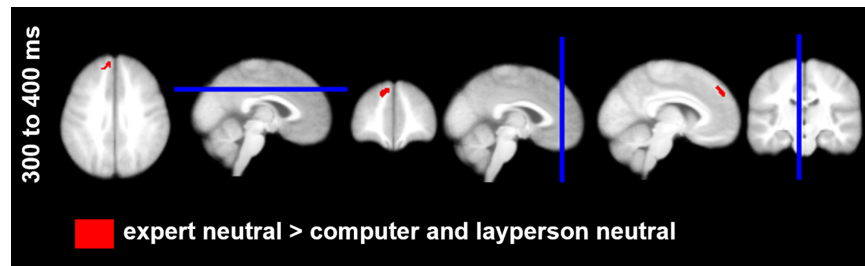**Fig. 6**. Source estimations for the main effect of emotion (post hoc t-tests are displayed, $P < .001$).

Fig. 7. Source estimations for sender by emotion interaction (post hoc *t*-tests are displayed, *P* < .001, within global sender by emotion interactions *P* < .001).

Table 3. Source estimations for emotion main effects

Cluster-level	Peak-level		MNI coordinates			LONI
Number of significant voxels	Peak t (1, 261)	Peak P-unc	x (mm)	y (mm)	z (mm)	area
P3 time window (300–400 ms)						
238 (60ᵃ)	4.91	<0.001	−32	−52	40	Sup parietal G L
213 (47ᵃ)	4.85	<0.001	−36	−50	36	Angular G L
977 (43ᵃ)	4.80	<0.001	6	−82	−8	Lingual G R
209	4.53	<0.001	48	−78	−6	Mid occipital G R
171	4.32	<0.001	0	−32	24	Cingulate G R
328	3.92	<0.001	−28	−84	10	Mid occipital G L
446	3.84	<0.001	−12	−78	−12	Lingual G L
36	3.30	<0.001	−22	2	64	Sup frontal G L
Early LPP time window (400–650 ms)						
483	4.22	<0.001	12	−84	−8	Lingual G R
108	4.09	<0.001	−30	−52	38	Sup parietal G L
89	4.03	<0.001	30	−50	38	Sup parietal G R
186	3.62	<0.001	−42	36	14	Inf frontal G L
145	3.33	<0.001	−12	−80	−12	Inf occipital G L
Late LPP time window (650–900 ms)						
102	3.63	<0.001	18	−88	−12	Inf occipital G R
75	3.42	<0.001	−10	−84	40	Sup occipital G L
60	3.27	<0.001	−38	34	14	Inf frontal G L

Notes. ᵃResulting cluster size when FWE-corrected threshold of *P* < .05 (≥25 significant voxels) was used. Number of significant voxels = the number of voxels that differ significantly between both conditions. Peak *P*-unc = uncorrected *P* value. For each significant peak, respective coordinates (*x*, *y* and *z*) are displayed in MNI space. A cluster may exhibit more than one peak, in which case only the highest peak is reported. Area = peak-level brain region as identified by the LONI atlas. R/L = laterality right or left. G = gyrus; Inf = inferior, Mid = middle, Sup = superior

Table 4. Source estimations for the sender by emotion interaction

Cluster-level	Peak-level		MNI coordinates			LONI
Number of significant voxels	Peak t (1, 261)	Peak P-unc	x (mm)	y (mm)	z (mm)	Area
P3 time window (300–400 ms)						
Expert neutral > layperson neutral						
62	3.85	<0.001	−6	52	32	Sup frontal G L
Expert neutral > computer neutral						
62	2.97	<0.005	−8	48	38	Sup frontal G L

Notes. Peak *P*-unc = uncorrected *P* value. For each significant peak, respective coordinates (*x*, *y* and *z*) are displayed in MNI space. Area = peak-level brain region as identified by the LONI atlas. R/L = laterality right or left. G = gyrus; Sup = superior

isolation of the effect of sender attribution in the absence of physical clues and previous interaction experience.

Attributed expertise rapidly drove brain responses in a linear fashion: within 200 ms, expert feedback prompted larger responses than did computer feedback, with layperson feedback intermediate. The general pattern persisted across all analyzed time windows but was more pronounced for early ERPs (N1, P2 and EPN). Here, a linear effect of expertise was clearly confirmed and corresponded to participants' ratings of feedback appropriateness. In line with the notion of a social prediction mechanism, responses to (on average) the same feedback were larger and occurred faster, when participants ascribed more expertise to the sender. Acceleration may be unique to human expertise, as it did not occur when more social competence was ascribed to a computer (Schindler & Kissler, 2016a). Scalp ERPs are consistent with attentional highlighting of feedback from apparently more significant senders and the general literature on object-based attention mechanisms (e.g. Ferrari *et al.*, 2008; Schoenfeld *et al.*, 2014), but source localization further implicates higher-level processes. The effect of affirmative expert feedback

in the N1/P2 time window was manifest in superior frontal and premotor activations in regions previously associated with self-reflection (Northoff et al., 2006; D'Argembeau et al., 2007; Akitsuki & Decety, 2009) and embodied processing (Agnew et al., 2007) and only later, in the EPN and P3 window, spread across occipital, parietal and additional frontal brain regions. The current temporal precedence of frontal over posterior brain activity might be due to the block design, where putative sender identity was kept constant within a block, enabling participants to prepare for significant feedback and potentially simulate what their presumed counterpart thought about them and was trying to convey to them.

Early emotion effects or interactions between emotion and sender were not detected. Although many single-word processing studies show such early emotion effects (e.g. Scott et al., 2009; Bayer et al., 2012), probably most robustly for the EPN (see Hinojosa et al., 2019, for review), previous studies using emotional adjectives in a feedback context have revealed more variability (Schindler et al., 2015; Schindler & Kissler, 2016a, 2018). This may be due to the pre-viewing phase in this as well as most of our previous feedback studies. Because there is time to process the word content during this pre-viewing phase, feedback disclosure may have foregrounded sender identity more than the content itself. Indeed, a separate analysis of the present pre-viewing phase revealed content main effects for both the N1 and the EPN (Schindler et al., 2019b). On the other hand, a different design using the same material as here, but disclosing the sender after word content, revealed late emotion effects, both during the pre-viewing period and during the feedback phase, suggesting a general temporal precedence of sender processing over content processing (Schindler & Kissler, 2018). Together, these data suggest that there is no fixed time point at which emotional significance is extracted from a word (and possibly also other stimuli) and that even minimal contexts can affect timing, although in typical psycho-linguistic experiments the EPN window seems most robustly responsive (see also Schindler et al., 2019b, for further discussion).

Enhanced processing of feedback from both human senders continued in the P3 window, where positive and negative feedback prompted larger responses than neutral one. Moreover, an intriguing interaction occurred: Neutral feedback from the expert elicited a higher P3 amplitude than the same feedback from the layperson, whereas for positive and negative feedback, amplitudes did not differ between expert and layperson.

Larger brain responses to neutral feedback from the expert may indicate a unique effect of human expertise and perhaps 'psychotherapy expertise' in particular: Weizenbaum (1966) noted that conversation partners tend to attribute deeper meaning to utterances from psychotherapists, even to those that may appear uninformed or non-sensical otherwise. He stated that this process 'has a crucial psychological utility in that it serves the speaker to maintain his sense of being heard and understood. The speaker further defends his impression (which in real life may be illusory), by attributing to his conversational partner all sorts of background knowledge, insights and reasoning ability. But again, these are the speaker's (i.e. Eliza program user's) contribution to the conversation' (Weizenbaum, 1966, p. 42). Regarding emotional feedback, specific expertise or authority seemed to play less of a role, as the responses to the putative layperson feedback did not differ from those to the expert. Only when the sender was denoted as acting randomly, suggesting no particular expertise, were brain responses to emotional feedback much smaller, although still significantly larger than those elicited by neutral feedback in this sender condition.

Supporting the notion that in particular neutral feedback from the expert prompts further cognitive processes in the receiver, source analysis revealed superior frontal cortex structures close to the mid-line as generators of the stronger response to neutral feedback supposedly given by the expert. These structures form part of a mentalizing network supporting higher-order social inferences from stories (Hervé et al., 2013) or complex scenes (Sugiura Motoaki et al., 2009). Similar activations have been observed as correlates of mentalizing about interactions with humans rather than computers (Kircher et al., 2009; Chaminade et al., 2012; Schindler, Kruse, et al., 2019a). Broad superior frontal networks close to the cerebral midline are also crucially involved in detecting and incorporating conflicting feedback from others (Welborn et al., 2016), eventually altering self-evaluations (Korn et al., 2012; Korn et al., 2016). Its present localization is therefore consistent with the thesis that participants' higher significance expectations for the more competent sender induce more mentalizing about inherently neutral feedback.

As a caveat regarding the above interpretation, we acknowledge that, at least regarding the rated person-descriptiveness, neutral adjectives were less concrete than positive and negative adjectives. Since concreteness and emotional content have been shown to interact, possible effects of this difference merit discussion: Whereas Kanske & Kotz (2007) and Palazova et al. (2013) reported bigger emotional–neutral ERP differences for concrete words, other studies report larger emotion effects for abstract words (Hinojosa et al., 2014; see also Hinojosa et al., 2019, for review). As these findings are somewhat inconsistent, it is difficult to predict how the fact that neutral traits had been rated as less concrete than emotional traits regarding their person-descriptiveness might have impacted the present findings. Still, we find it unlikely that an interaction between emotion and concreteness would affect only one specific sender block (expert) or only one specific ERP component or that the effect would localize to a brain region that is typically and most prominently associated with theory of mind, mentalizing and social cognition (see neurosynth.org database for the respective top associations). Nevertheless, this possibility needs to be considered. We hope to settle the issue empirically in the future.

For the LPP, separate main effects of sender and emotion revealed a stronger impact of human feedback in general as well as of positive and negative feedback, replicating previous research (Schindler et al., 2015; Schindler & Kissler, 2016, 2018) and further confirming larger late ERPs for emotional than for neutral contents (Kissler et al., 2009; Schacht & Sommer, 2009a; Hinojosa et al., 2010).

LPP main effects were localized in the extended visual cortex, where sender and content effects partly overlapped, as well as in parietal, somatosensory and frontal areas. Visual and superior occipital activations are typically found in reading studies (Osipowicz et al., 2011; van der Mark et al., 2011), and the present pattern is in line with emotional content (Vuilleumier, 2005; Schupp et al., 2006; Pessoa & Adolphs, 2010) and communicative context tuning mechanisms of 'motivated attention' (Lang et al., 1997), confirming its extension from stimulus-driven emotional to contextual social significance (see also Schindler et al., 2015).

Furthermore, posterior cingulate regions were found to be activated by putative human feedback. The PCC, next to visual and lingual areas, shows reliable activation in response to written narratives (Regev et al., 2013) and has been likewise suggested to sub-serve controlled and automatic self-reflective processes (Lieberman, 2007). Supporting evidence for the PCC's role also comes from social preference tasks (Chen et al., 2010) or com-

parisons of evaluative feedback with performance feedback (Pan *et al.,* 2009). Therefore, the PCC along with the medial prefrontal cortex is seen as an integral node of the metalizing network (Northoff & Bermpohl, 2004; Uddin *et al.,* 2007; Schilbach *et al.,* 2012). The fact that PCC, next to frontal and somatosensory activations, was also found active in a recent fMRI variant of this paradigm (Schindler and Kissler, 2018) adds to the credibility of these localizations, although PCC is a relatively deep structure and has not been localized in the previous EEG studies (Schindler *et al.,* 2015; Schindler & Kissler, 2016a). Increased experimental power as well as the increased complexity and mentalizing requirements of the present scenario may have contributed to this finding.

Sender attributions activated frontal areas earlier and to a greater extent than did emotional content. Premotor and somatosensory activations are consistent with the notion of embodied processing as an integral part of both social cognition and processing of emotional stimuli (Niedenthal *et al.,* 2005). The time course of the effects suggests that parsing of sender identity, as an important social context variable, sets the stage for subsequent content processing which it can dynamically tune. Notably, the extent to which structures theoretically related to somatosensory processing were activated by identical stimuli depended on the perceived feedback source, implying that embodied processing, that is the (co-)activation of brain structures that represent our bodies and mediate awareness of bodily states, is not a fixed property of stimuli such as emotion concepts (Niedenthal *et al.,* 2009).

Both feedback source, specifically attributed sender competence and feedback content activated visual cortex for several hundred milliseconds. Yet, these effects showed distinct temporal profiles, with source effects preceding content effects and increasing mostly linearly with attributed sender competence, whereas content effects started only with the P3 and mostly responded in a u-shaped manner to emotional content, essentially reflecting stimulus arousal. On the LPP, both attributed feedback source and content acted in parallel, raising the question of what drives visual cortex activity here. For one, intrinsic properties of the visual cortex may exhibit such different tuning profiles: sender and content effects, while both increasing visual cortex activity, may still be separable within the extended visual cortex via multivariate methods (e.g. Contini *et al.,* 2017; King & Deheane, 2014). On the other hand, regarding similar effects of emotion and attention on the visual cortex (e.g. Ferrari, Cardinale, & Codispoti, 2008), parallel re-entrant amplification effects from distant cerebral structures have been implicated as drivers of visual emotion effects. This is most notably the case for projections originating in the amygdala (e.g. Vuilleumier, 2005). Similar principles may well be at work in the present scenario. Some of the frontal, parietal or even cingulate sources localized for the sender effects may drive the visual cortex. In parallel, subcortical structures too deep to be localized in an EEG study may also be active. In fact, a simpler fMRI variant of this paradigm revealed subcortical activations in the amygdala and ventral striatum for feedback content and in extended mentalizing regions for feedback source (Schindler et al., 2019). Whether and what kind of functional relationship exists between the visual cortex and other cerebral regions in the present paradigm awaits further investigation.

Overall, present results indicate that structures involved in mentalizing and embodied simulation are rapidly activated to determine the personal significance of single-word feedback in a quasi-interactive scenario, underscoring that subjective significance is an emergent property of processing the entire communicative situation as it is represented by the individual (e.g. Blumer, 1969). In the age of online virtual interaction, communication often relies on representations that build on few, if any, physical cues, underscoring the importance of understanding the attribution mechanisms involved. These play a key role in online therapy and may also underlie manipulation of public opinion via electronic media. Combining EEG's inherently high temporal resolution with advanced source analysis methods afforded by high-density scalp coverage, this study delineates some of these mechanisms.

Supplementary data

Supplementary data are available at *SCAN* online.

Acknowledgements

References

Agnew, Z.K., Bhakoo, K.K., Puri, B.K. (2007). The human mirror system: a motor resonance theory of mind-reading. *Brain Research Reviews*, **54**, 286–93.

Aguado, L., Parkington, K.B., Dieguez-Risco, T., *et al.* (2019). Joint modulation of facial expression processing by contextual congruency and task demands. *Brain Sciences*, **9**, 116.

Akitsuki, Y., Decety, J. (2009). Social context and perceived agency affects empathy for pain: an event-related fMRI investigation. *NeuroImage*, **47**, 722–34.

Motoaki, S., Keisuke, W., Atsushi, S., *et al.* (2009). Extraction of situational meaning by integrating multiple meanings in a complex environment: a functional MRI study. *Human Brain Mapping*, **30**, 2676–88.

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, **11**, 280–9.

Barak, A., Klein, B., Proudfoot, J.G. (2009). Defining internet-supported therapeutic interventions. *Annals of Behavioral Medicine*, **38**, 4–17.

Barrett, L.F., Lindquist, K.A., Gendron, M. (2007). Language as context for the perception of emotion. *Trends in Cognitive Sciences*, **11**, 327–32.

Barrett, L.F., Mesquita, B., Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science*, **20**, 286–90.

Bayer, M., Sommer, W., Schacht, A. (2012). Font size matters—emotion and attention in cortical responses to written words. *PLoS One*, **7**, e36042.

Bayer, M., Ruthmann, K., Schacht, A. (2017). The impact of personal relevance on emotion processing: evidence from event-related potentials and pupillary responses. *Social Cognitive and Affective Neuroscience*, **12**, 1470–9.

Beck, A.T., Steer, R.A., Hautzinger, M. (2001). *Beck-Depressions-Inventar (BDI): Testhandbuch. 2., überarb. Aufl., 1. Nachdr*, Bern: Huber.

Blumer, H. (1969). *Symbolic Interactionism: Perspective & Method*, New York: Prentice Hall.

Chaminade, T., Rosset, D., Fonseca, D.D., *et al.* (2012). How do we think machines think? An fMRI study of alleged competition with an artificial intelligence. *Frontiers in Human Neuroscience*, **6**.

Chen, A.C., Welsh, R.C., Liberzon, I., *et al.* (2010). 'Do I like this person?' A network analysis of midline cortex during a social preference task. *NeuroImage*, **51**, 930–9.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn, Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Collins, D.R., Stukas, A.A. (2006). The effects of feedback self-consistency, therapist status, and attitude toward therapy on reaction to personality feedback. *The Journal of Social Psychology*, **146**, 463–83.

Contini, E.W., Wardle, S.G., Carlson, T.A. (2017). Decoding the time-course of object recognition in the human brain: from visual features to categorical decisions. *Neuropsychologia*, **105**, 165–76.

D'Argembeau, A., Ruby, P., Collette, F., *et al.* (2007). Distinct regions of the medial prefrontal cortex are associated with self-referential processing and perspective taking. *Journal of Cognitive Neuroscience*, **19**, 935–44.

Diéguez-Risco, T., Aguado, L., Albert, J., *et al.* (2013). Faces in context: modulation of expression processing by situational information. *Social Neuroscience*, **8**, 601–20.

Diéguez-Risco, T., Aguado, L., Albert, J., *et al.* (2015). Judging emotional congruency: explicit attention to situational context modulates processing of facial expressions of emotion. *Biological Psychology*, **112**, 27–38.

Ferrara, E., Varol, O., Davis, C., *et al.* (2016). The rise of social bots. *Communications of the ACM*, **59**, 96–104.

Ferrari, V., Codispoti, M., Cardinale, R., *et al.* (2008). Directed and motivated attention during processing of natural scenes. *Journal of Cognitive Neuroscience*, **20**, 1753–61.

Fields, E.C., Kuperberg, G.R. (2012). It's all about you: an ERP study of emotion and self-relevance in discourse. *NeuroImage*, **62**, 562–74.

Fischler, I., Bradley, M. (2006). Event-related potential studies of language and emotion: words, phrases, and task effects. In: Anders, S., Ende, G., Junghofer, M., *et al.*, editors. *Progress in Brain Research. Understanding Emotions*, Elsevier, pp. 185–203.

Frenkel, T.I., Bar-Haim, Y. (2011). Neural activation during the processing of ambiguous fearful facial expressions: an ERP study in anxious and nonanxious individuals. *Biological Psychology*, **88**, 188–95.

Gazzola, V., Spezio, M.L., Etzel, J.A., *et al.* (2012). Primary somatosensory cortex discriminates affective significance in social touch. *Proceedings of the National Academy of Sciences*, **109**, E1657–66.

Gilbert, D.T., Wilson, T.D. (2007). Prospection: experiencing the future. *Science*, **317**, 1351–4.

Hajcak, G., Dunning, J.P., Foti, D. (2009). Motivated and controlled attention to emotion: time-course of the late positive potential. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, **120**, 505–10.

Herbert, C., Junghöfer, M., Kissler, J. (2008). Event related potentials to emotional adjectives during reading. *Psychophysiology*, **45**, 487–98.

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**, 653–61.

Hervé, P.-Y., Razafimandimby, A., Jobard, G., *et al.* (2013). A shared neural substrate for mentalizing and the affective component of sentence comprehension. *PLoS One*, **8**, e54400.

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**, 748–57.

Hinojosa, J.A., Albert, J., López-Martín, S., *et al.* (2014). Temporospatial analysis of explicit and implicit processing of negative content during word comprehension. *Brain and Cognition*, **87**, 109–21.

Hinojosa, J.A., Moreno, E.M., Ferré, P. (2019). Affective neurolinguistics: towards a framework for reconciling language and emotion. *Language, Cognition and Neuroscience*, 0, 1–27

Hofmann, M.J., Kuchinke, L., Tamm, S., *et al.* (2009). Affective processing within 1/10th of a second: high arousal is necessary for early facilitative processing of negative but not positive words. *Cognitive, Affective, & Behavioral Neuroscience*, **9**, 389–97.

Hughes, B.L., Beer, J.S. (2012). Protecting the self: the effect of social-evaluative threat on neural representations of self. *Journal of Cognitive Neuroscience*, **25**, 613–22.

Jacoby, N., Bruneau, E., Koster-Hale, J., *et al.* (2016). Localizing Pain Matrix and Theory of Mind networks with both verbal and non-verbal stimuli. *NeuroImage*, **126**, 39–48.

Kanske, P., Kotz, S.A. (2007). Concreteness in emotional words: ERP evidence from a hemifield study. *Brain Research*, **1148**, 138–48 Epub 2007 Feb 27.

Kanske, P., Plitschka, J., Kotz, S.A. (2011). Attentional orienting towards emotion: P2 and N400 ERP effects. *Neuropsychologia*, **49**, 3121–9.

Keuper, K., Zwitserlood, P., Rehbein, M.A., *et al.* (2013). Early prefrontal brain responses to the hedonic quality of emotional words–a simultaneous EEG and MEG study. *PLoS One*, **8**, e70788.

Keuper, K., Zwanzger, P., Nordt, M., *et al.* (2014). How 'love' and 'hate' differ from 'sleep': using combined electro/magnetoencephalographic data to reveal the sources of early cortical responses to emotional words. *Human Brain Mapping*, **35**, 875–88.

King, J.-R., Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in Cognitive Sciences*, **18**, 203–10.

Kircher, T., Blumel, I., Marjoram, D., *et al.* (2009). Online mentalising investigated with functional MRI. *Neuroscience Letters*, **454**, 176–81. doi: 10.1016/j.neulet.2009.03.026.

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

Kissler, J., Herbert, C., Winkler, I., *et al.* (2009). Emotion and attention in visual word processing - an ERP study. *Biological Psychology*, **80**, 75–83.

Klein, F., Iffland, B., Schindler, S., *et al.* (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, & Behavioral Neuroscience*, **15**, 736–48.

Korn, C.W., Prehn, K., Park, S.Q., *et al.* (2012). Positively biased processing of self-relevant social feedback. *The Journal of Neuroscience*, **32**, 16832–44.

Korn, C.W., Rosée, L.L., Heekeren, H.R., *et al.* (2016). Social feedback processing in borderline personality disorder. *Psychological Medicine*, **46**, 575–87.

Lang, P.J., Melamed, B.G., Hart, J. (1970). A psychophysiological analysis of fear modification using an automated desensitization procedure. *Journal of Abnormal Psychology*, **76**, 220–34.

Lang, P.J., Bradley, M.M., Cuthbert, B.N. (1997). Motivated attention: Affect, activation, and action). In: Lang, P.J., Simons, R.F., Balaban, M.T., editors. *Attention and Orienting: Sensory and Motivational Processes*, Hillsdale, NJ: Erlbaum, pp. 97–135.

Lee, P. (2016). Learning from Tay's introduction. In: *The Official Microsoft Blog* [online]. Available from: http://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/, [Accessed July 15, 2016].

Leuthold, H., Filik, R., Murphy, K., *et al.* (2012). The on-line processing of socio-emotional information in prototypical scenarios: inferences from brain potentials. *Social Cognitive and Affective Neuroscience*, **7**, 457–66.

Li, S., Zhu, X., Ding, R., *et al.* (2019). The effect of emotional and self-referential contexts on ERP responses towards surprised faces. *Biological Psychology*, 107728.

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

Litvak, V., Friston, K. (2008). Electromagnetic source reconstruction for group studies. *NeuroImage*, **42**, 1490–8. doi: 10.1016/j.neuroimage.2008.06.022.

Litvak, V., Mattout, J., Kiebel, S., *et al.* (2011). EEG and MEG data analysis in SPM8. *Computational Intelligence and Neuroscience*, **2011**, 1–32.

van der Mark, S., Klaver, P., Bucher, K., *et al.* (2011). The left occipitotemporal system in reading: disruption of focal fMRI connectivity to left inferior frontal and inferior parietal language areas in children with dyslexia. *NeuroImage*, **54**, 2426–36.

Niedenthal, P.M., Barsalou, L.W., Winkielman, P., *et al.* (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review*, **9**, 184–211.

Niedenthal, P.M., Winkielman, P., Mondillon, L., *et al.* (2009). Embodiment of emotion concepts. *Journal of Personality and Social Psychology*, **96**, 1120.

Northoff, G., Bermpohl, F. (2004). Cortical midline structures and the self. *Trends in Cognitive Sciences*, **8**, 102–7.

Northoff, G., Heinzel, A., de Greck, M., *et al.* (2006). Self-referential processing in our brain—a meta-analysis of imaging studies on the self. *NeuroImage*, **31**, 440–57

Osipowicz, K., Rickards, T., Shah, A., *et al.* (2011). A test of the role of the medial temporal lobe in single-word decoding. *NeuroImage*, **54**, 1455–64.

Palazova, M., Sommer, W., Schacht, A. (2013). Interplay of emotional valence and concreteness in word processing: an event-related potential study with verbs. *Brain and Language*, **125**, 264–71.

Pan, X., Hu, Y., Li, L., *et al.* (2009). Evaluative-feedback stimuli selectively activate the self-related brain area: an fMRI study. *Neuroscience Letters*, **465**, 90–4.

Pessoa, L., Adolphs, R. (2010). Emotion processing and the amygdala: from a 'low road' to 'many roads' of evaluating biological significance. *Nature Reviews Neuroscience*, **11**, 773–83.

Peyk, P., De Cesarei, A., Junghöfer, M. (2011, 2011, Article ID). Electro Magneto Encephalograhy Software: overview and integration with other EEG/MEG toolboxes. *Computational Intelligence and Neuroscience*, **861705**, 1–11.

Regev, M., Honey, C.J., Simony, E., *et al.* (2013). Selective and invariant neural responses to spoken and written narratives. *The Journal of Neuroscience*, **33**, 15978–88.

Rohr, L., Abdel Rahman, R. (2015). Affective responses to emotional words are boosted in communicative situations. *NeuroImage*, **109**, 273–82.

Sass, S.M., Heller, W., Stewart, J.L., *et al.* (2010). Time course of attentional bias in anxiety: emotion and gender specificity. *Psychophysiology*, **47**, 247–59.

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

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

Schilbach, L., Bzdok, D., Timmermans, B., *et al.* (2012). Introspective minds: using ALE meta-analyses to study commonalities in the neural correlates of emotional processing, social & unconstrained cognition. *PloS One*, **7**, e30920–0.

Schindler, S., Kissler, J. (2016a). People matter: perceived sender identity modulates cerebral processing of socio-emotional language feedback. *NeuroImage*, **134**, 160–9.

Schindler, S., Kissler, J. (2016b). Selective visual attention to emotional words: early parallel frontal and visual activations followed by interactive effects in visual cortex. *Human Brain Mapping*, **37**, 3575–87.

Schindler, S., Kissler, J. (2018). Language-based social feedback processing with randomized "senders": an ERP study. *Social Neuroscience*, **13**, 202–13.

Schindler, S., Wegrzyn, M., Steppacher, I., *et al.* (2015). Perceived communicative context and emotional content amplify visual word processing in the fusiform gyrus. *The Journal of Neuroscience*, **35**, 6010–9.

Schindler, S., Kruse, O., Stark, R., *et al.* (2019a). Attributed social context and emotional content recruit frontal and limbic brain regions during virtual feedback processing. *Cognitive, Affective, & Behavioral Neuroscience*, **19**, 239–52.

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

Schoenfeld, M.A., Hopf, J.-M., Merkel, C., *et al.* (2014). Object-based attention involves the sequential activation of feature-specific cortical modules. *Nature Neuroscience*, **17**, 619–24.

Schupp, H.T., Öhman, A., Junghöfer, M., *et al.* (2004). The facilitated processing of threatening faces: an ERP analysis. *Emotion*, **4**, 189.

Schupp, H.T., Flaisch, T., Stockburger, J., *et al.* (2006). Emotion and attention: event-related brain potential studies. *Progress in Brain Research*, **156**, 31–51.

Schupp, H.T., Stockburger, J., Codispoti, M., *et al.* (2007). Selective visual attention to emotion. *The Journal of Neuroscience*, **27**, 1082–9.

Scott, G.G., O'Donnell, P.J., Leuthold, H., *et al.* (2009). Early emotion word processing: evidence from event-related potentials. *Biological Psychology*, **80**, 95–104. doi: 10.1016 2008.03.010 Epub 2008 Mar 22.

Shattuck, D.W., Mirza, M., Adisetiyo, V., *et al.* (2008). Construction of a 3D probabilistic atlas of human cortical structures. *NeuroImage*, **39**, 1064–80.

Shrauger, J.S., Schoeneman, T.J. (1979). Symbolic interactionist view of self-concept: through the looking glass darkly. *Psychological Bulletin*, **86**, 549.

Spielberger, C.D., Sydeman, S.J., Owen, A.E., *et al.* (1999). Measuring anxiety and anger with the State-Trait Anxiety Inventory (STAI) and the State-Trait Anger Expression Inventory (STAXI). In: Maruish, M.E., editor. *The Use of Psychological Testing for Treatment Planning and Outcomes Assessment*, 2nd edn, Lawrence Erlbaum Associates: Mahwah, pp. 993–1021.

Uddin, L.Q., Iacoboni, M., Lange, C., *et al.* (2007). The self and social cognition: the role of cortical midline structures and mirror neurons. *Trends in Cognitive Sciences*, **11**, 153–7.

Vuilleumier, P. (2005). How brains beware: neural mechanisms of emotional attention. *Trends in Cognitive Sciences*, **9**, 585–94.

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, **9**, 36–45.

Welborn, B.L., Lieberman, M.D., Goldenberg, D., *et al*. (2016). Neural mechanisms of social influence in adolescence. *Social Cognitive and Affective Neuroscience*, **11**, 100–9.

Wieser, M.J., Gerdes, A.B.M., Büngel, I., *et al*. (2014). Not so harmless anymore: how context impacts the perception and electrocortical processing of neutral faces. *NeuroImage*, **92**, 74–82.