# Do people get used to insulting language?

Struiksma, M.E.; De Mulder, H.N.M.; Van Berkum, J.J.A.

# Do People Get Used to Insulting Language?

Marijn E. Struiksma[1]*, Hannah N. M. De Mulder[2] and Jos J. A. Van Berkum[1]

[1] Department of Languages, Literature and Communication, UiL OTS/Institute for Language Sciences, Utrecht University, Utrecht, Netherlands, [2] Leiden University Centre for Linguistics (LUCL), Leiden University, Leiden, Netherlands

Verbal insults go against a universal moral imperative not to inflict harm on others, and they also pose a threat to one's face or reputation. As such, these "verbal slaps in the face" provide a unique opportunity to explore the interface between language and emotion. We used electroencephalography (EEG) and skin conductance (SC) recordings to compare the short-term impact of verbal insults such as "Linda is an idiot" or "Paula is horrible" to that of more positive evaluations (e.g., "Linda is an angel", "Paula is impressive") and neutral factual descriptions (e.g., "Linda is a student"), examining how responses to these various speech acts adapt as a function of massive repetition. By using either the participant's or somebody else's name, we also explored how statement impact depended on who was being evaluated. Multilevel ERP analysis with three predesignated latency ranges revealed an early insult effect in P2 amplitude that was highly robust over repetition and also did not depend on who the insult was about. This P2 effect points to a very rapid and stable capture of emotional attention, plausibly triggered by the retrieval of evaluative word meaning from long-term memory. Insults also elicited a larger late positive potential (LPP), again regardless of who the insult was about, but this later effect did not withstand repetition. Skin conductance responses showed that insults did not lead to more arousal than compliments did. In all, our findings suggest that in a standard psycholinguistic comprehension experiment without real interaction between speakers, insults deliver lexical "mini-slaps in the face," such that the strongly negative evaluative words involved (e.g., "idiot") automatically grab attention during lexical retrieval, regardless of how often that retrieval occurs.

Keywords: psycholinguistics, communication, emotion, insults, morality, EEG, ERP, skin conductance

## INTRODUCTION

"Sticks and stones may break my bones, but words will never hurt me." Although consoling as an idea, we all know that this is not really true. Words can hurt, for at least one obvious reason: they are used to realize interpersonal behavior, and interpersonal behavior can sometimes really hurt. Think about receiving a break-up message, being criticized or insulted, or mocked. What is not so obvious, though, is the exact way in which words deliver their offensive, emotionally negative payload at the moment these words are being read or heard. Language comprehension is a highly complex multi-faceted process that rapidly activates or generates a variety of representations at various moments in time (e.g., Jackendoff, 2007; Tomasello, 2008; Van Berkum, 2018, 2019), such as the meaning of each of the words as retrieved from memory, the situation these words refer to, and the social intentions that can be attributed to the speaker.

This cascade of computed representations opens up the possibility that words can actually be offensive in multiple ways, related to different steps in the cascade, and at different moments in time.

Here, we explore the interface between words and their effects on people by examining the short-term impact of a particular type of offensive language, the verbal insult. One of the reasons why offensive expressions such as "you are ugly" or "you're a real asshole!" are frowned upon is because they violate a universal moral imperative to not inflict harm on others (Haidt, 2012; Greene, 2013; Gantman and Van Bavel, 2015). Despite this moral imperative, however, insults occur rather frequently in everyday life, in all corners of society, and with potentially serious consequences (Jay, 2009; Irvine, 2013). Understanding what an insulting expression does to people as it unfolds, and why, is therefore of considerable importance—to psycholinguists interested in how language moves people, but also to others who wish to understand the details of social behavior.

Everyday experience suggests that insults behave differently from more benign speech acts not only in their more negative "payload", but also in how easily they wear off. There are at least two sides to this. First, and sadly, many people seem to find it quite easy to brush off a compliment and forget about it as the day progresses, whereas a single insult may bother us for the rest of the day, and beyond. Second, whereas the impact of repeated compliments seems to adapt with repetition (just as many other good things do) or can in some cases even have detrimental effects (e.g., Brummelman et al., 2017), repeated insults do not seem to lose their sting. Of course, these are mostly informal observations. The main goal of our study is to examine whether evidence for the differential adaptation (or "saturation") to repeated verbal insults and repeated compliments can be observed in the lab, and, if so, which stage(s) of processing are implicated in the adaptation, and which are not. Such evidence can not only help us work out how insults and other moral transgressions are processed, initially and when repeated, but can also more generally advance our understanding of the complex interface between language and emotion. In passing, we may gain more insight into the everyday experience of why some things continue to move us, while others do not.

One relevant dimension is whether *you* or somebody else is insulted. Insults directed at you pose a severe threat to the self as well as to your reputation, and might for that reason not easily lose their sting—for members of an ultrasocial species that specializes in cooperation beyond the family (Leary, 2005; Tennie et al., 2010; Tomasello and Herrmann, 2010; Baumeister and Bushman, 2021), threats to one's reputation are not to be taken lightly. On the other hand, insults might also resist adaptation regardless of who the person evaluated is, because they inflict harm on others, because they are informative as to who is willing to do so, and because they signal a social conflict in your vicinity, possibly even in your group. Members of an ultrasocial species may well want to pay attention to such nearby verbal "slaps in the face" (Irvine, 2013).

On the reasonable assumption that verbal insults trigger a cascade of rapidly consecutive or overlapping processing effects, different parts of that cascade might be differently affected by repetition, with some of them rapidly wearing off, and others remaining strongly responsive for quite some time. In this study, we therefore use electroencephalography (EEG), a method that provides high temporal resolution as well as some qualitative indications of the processes involved.

## Processing Nonverbal Emotional Stimuli

Emotional pictures and sounds attract attention and, as such, are given priority in subsequent processing (e.g., Vuilleumier, 2005; Schupp et al., 2006; Frijda, 2007; Kissler et al., 2009; Damasio, 2010; Panksepp and Biven, 2012; Pessoa, 2013; Carretié, 2014; Schupp and Kirmse, 2021). Many EEG studies have revealed such emotion-induced attention capture. In EEG research with emotional and neutral pictures, for example, emotional pictures can increase the P2, an anteriorly distributed short-lived ERP component around 200–250 ms after picture onset (Carretié et al., 2001; Delplanque et al., 2004; Huang and Luo, 2006). Emotional pictures have also been shown to modulate somewhat later EEG components, such as the early posterior negativity or EPN around 250–300 ms (Schupp et al., 2004, 2006; Herbert et al., 2008; Langeslag and Van Strien, 2018; Frank and Sabatinelli, 2019; Schupp and Kirmse, 2021), and the late positive potential or LPP, a deflection that typically emerges somewhere after 300 ms over large areas of the scalp, with a centro-parietal maximum, and a possibly lengthy duration (Schupp et al., 2004; Hajcak et al., 2013; Sabatinelli et al., 2013; Frank and Sabatinelli, 2019; Hajcak and Foti, 2020; Schupp and Kirmse, 2021).

As can be expected, negative and positive emotional pictures and sounds capture attention and are given priority in subsequent processing (Carretié, 2014). However, research on the negativity bias has revealed that people are on average particularly sensitive to negative events: such events not only recruit more attention and intensified processing than neutral events, but also often do so relative to positive events (Ito et al., 1998; Carretié et al., 2001; Rozin and Royzman, 2001; Huang and Luo, 2006; see Carretié, 2014, for a review). In EEG studies, for example, negative pictures have been reported to elicit a larger P2 (Carretié et al., 2001; Huang and Luo, 2006), a larger EPN (Schupp et al., 2004, 2006; Herbert et al., 2008), and a larger LPP (Ito et al., 1998; Huang and Luo, 2006; Foti et al., 2009) as compared to positive pictures. The exact source of the bias is currently under debate, with some arguing that it simply reflects statistical properties of the environment (Shin and Niv, 2021; Unkelbach et al., 2021), and others proposing an evolutionary analysis involving the degree to which negative vs. positive stimuli affect fitness (Lazarus, 2021). Also, the negativity bias does not guarantee that *every* negative stimulus or stimulus set captures more attention than every positive stimulus or stimulus set (Carretié, 2014). After all, a snapping shoelace is a lot less evocative than the birth of one's child. The negativity bias is real, but it exists as an average phenomenon, emerging for reasons that remain to be fully explained.

## Processing Emotional Language

As might be expected, similar mechanisms of attention capture and subsequent intensified processing are at work when people read or listen to emotional language (e.g., Bernat et al., 2001;

Kissler et al., 2006, 2009; Kanske and Kotz, 2007; Herbert et al., 2008; Scott et al., 2009; Lei et al., 2017; see Citron, 2012 for a review). For example, relative to neutral words, negative (and positive) words have been reported to elicit an enhanced P2 (Herbert et al., 2006; Trauer et al., 2012; Wang and Bastiaansen, 2014), an enhanced EPN (e.g., Herbert et al., 2008; Recio et al., 2014; Rohr and Abdel Rahman, 2015), an enhanced N400 (e.g., Kanske and Kotz, 2007; Rohr and Abdel Rahman, 2015, for negative words only), as well as a larger LPP (Carretié et al., 2008; Herbert et al., 2008; Kissler et al., 2009; Wang and Bastiaansen, 2014).[1] Comparable emotion-induced EEG effects can be observed for words that are part of a larger discourse (e.g., Holt et al., 2009; Fields and Kuperberg, 2012, 2015, 2016). Although comparisons of emotional to neutral words often involve both valence and the degree of arousal (sympathetic system activation) that the word brings about, there is evidence that either one can play a role in the effects obtained, and that the two factors can also interact (see, e.g., Bayer et al., 2010; Citron et al., 2013; Recio et al., 2014; Espuny et al., 2018; Vieitez et al., 2021).

In addition to being a negative event, a verbal insult is also a morally objectionable statement, which makes EEG studies on the processing of morally objectionable language relevant to current concerns. In an early study exploring the impact of morally objectionable language with ERPs (Van Berkum et al., 2009), participants read statements that did or did not clash with their moral value system. Relative to acceptable words, words that rendered the unfolding statement morally objectionable for the participant at hand elicited an early P2-like positivity around 200-250 ms, and an increased LPP between 500 and 650 ms. These words also elicited a small centro-parietally distributed N400 effect, a well-known ERP effect that within the language domain is assumed to index some context-induced difficulty in retrieving and/or otherwise processing the meaning of a word (e.g., Kutas and Federmeier, 2000, 2011; Van Berkum, 2009; Brouwer et al., 2017). These various ERP effects have also been observed in other moral transgression studies (Foucart et al., 2015; Leuthold et al., 2015; Peng et al., 2017; Kunkel et al., 2018; Weimer et al., 2019), albeit not always all together. To the extent that they reappear in the current study, they can be useful in tracking the impact of a verbal transgression as a function of repeating it.

A few EEG studies have directly examined the processing of insulting language. In studies using isolated words, insulting words have been reported to elicit LPP effects (Carretié et al., 2008, relative to neutral but not complimenting words), occipital P1 effects (Wabnitz et al., 2012, relative to both), and increased mid-latency frontal negativities (Wabnitz et al., 2012, between 450 and 580 ms relative to neutral but not complimenting words; Wabnitz et al., 2016, between 395-480 ms, relative to both). EEG research on richer and more realistic insults is actually very rare, but in two recent studies with somewhat more contextualized

insults, insulting words elicited a centro-parietally focused LPP effect between 500 and 800 ms in one study (Rohr and Abdel Rahman, 2018, relative to neutral but not complimenting words), and an increased anterior negativity between 300 and 400 ms followed by a small LPP effect between 600 and 900 ms in the other study (Otten et al., 2017, relative to compliments).

Although relative to compliments, insults do not always increase the amplitude of potentially relevant components, such as the P1, P2, EPN, N400 and LPP (e.g., Benau and Atchley, 2020) and certainly do not boost these components simultaneously in every study, the relevant EEG studies conducted so far do suggest that insulting language does typically capture attention, and elicits intensified processing very rapidly, in a way that is comparable to the impact of other verbal moral transgressions (e.g., Van Berkum et al., 2009). Behavioral research on verbal insults confirms this. In a lexical decision task, for example, word-pseudoword decisions are slower and more error-prone with insult words than with compliment words (Carretié et al., 2008). Furthermore, in emotional Stroop tasks, people are slower to name the color of insult words than of compliment words (Siakaluk et al., 2011). In all, and as might be expected from the negativity bias literature, verbal insults command attention to a larger extent than verbal compliments do, presumably due to their offensive nature.

## What Exactly Is Offensive About an Insult?

What aspects of a verbal insult can give rise to such effects? Imagine hearing "you are an idiot!", uttered toward you in a way that is genuinely contemptuous. A recent model of affective language comprehension, the ALC model (Van Berkum, 2018, 2019; see **Figure 1**), predicts that multiple aspects of this verbal event can have an emotional impact. Most obviously, the speaker realizes a social intention (cf. Tomasello, 2008; Enfield, 2013), which in this case presumably is to overtly express his or her contempt for you, or at least to strike out at you, and possibly cause some hurt. That social act or "move" will usually elicit strong emotion—as part of our evolved cooperation-oriented way of living, human beings are all striving for respect (MacDonald and Leary, 2005; Greene, 2013), and an insult is a serious threat to one's "face" (Goffman, 1967; Brown and Levinson, 1987). Related to this, people want to "belong" and be close to at least some others, and a genuine insult signals strong interpersonal distancing—depending on who is speaking, that might hurt a bit, or a lot. At the level of social moves, genuine insults are very unpleasant things to receive.

The ALC model, however, predicts that an unfolding utterance such as "you are an idiot!" can generate at least three other emotionally evocative representations, that, although related to the social move, are not identical to it. One is the inferred referential intention of the speaker (cf. Tomasello, 2008): while the situation described by the speaker (you being an idiot) need not be a correct characterization of the state of affairs in the world, imagining this state as an unavoidable part of processing the utterance may nevertheless still elicit some emotion. A second possible trigger is the speaker's stance: a contemptuously uttered "you are an idiot!" will inevitably lead the listener to register contempt, regardless of what the exact target of that contempt

---

[1]The relationship of the LPP to various other late and roughly similarly distributed positive ERP deflections, such as the P3, P600, and Late Positive Complex (LPC) is currently under debate (e.g., Hajcak and Foti, 2020). Starting from emotion-oriented ERP research, we focus on LPP-relevant studies here, but we'll briefly return to the relationship with the P3 in the discussion section.
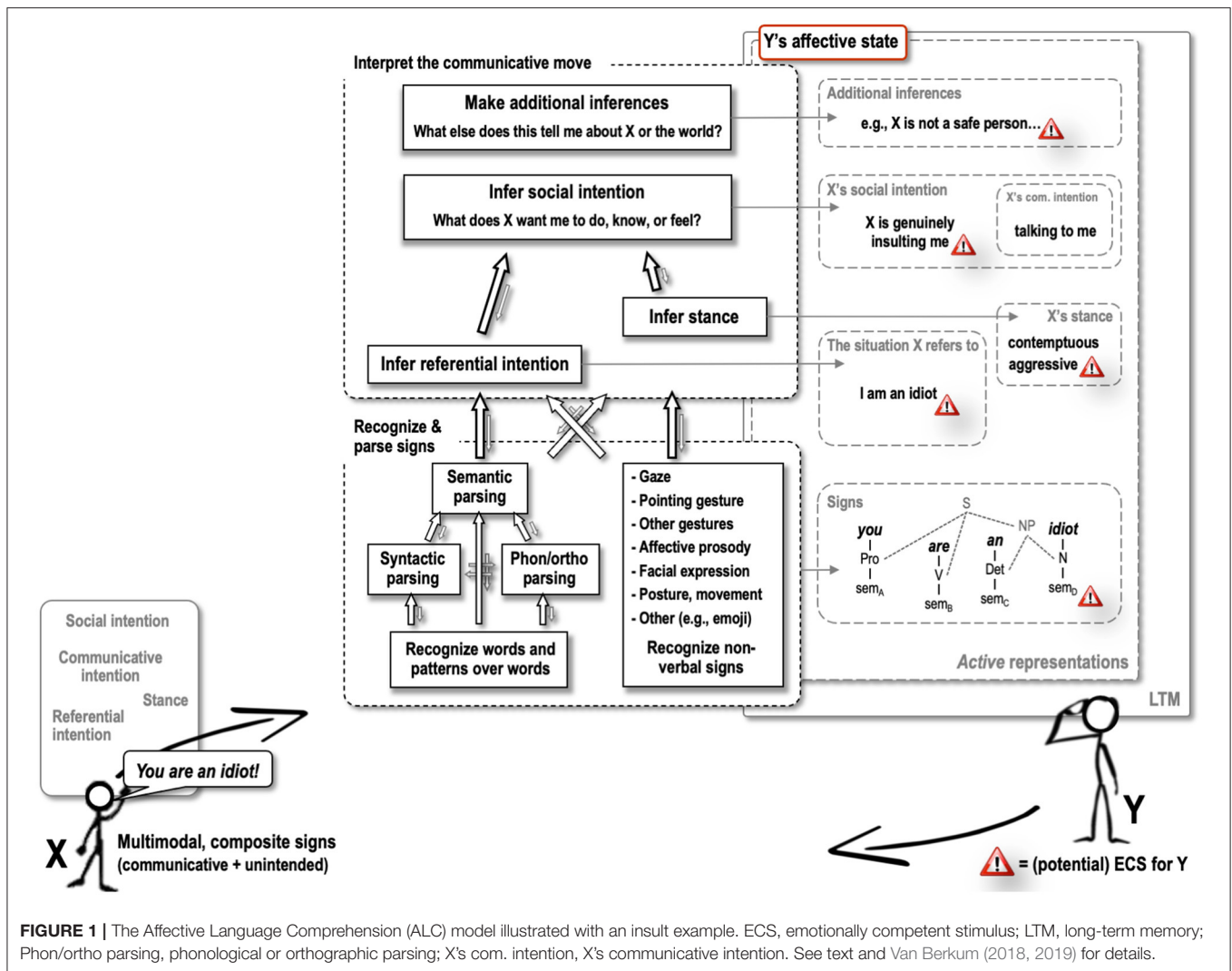
**FIGURE 1 |** The Affective Language Comprehension (ALC) model illustrated with an insult example. ECS, emotionally competent stimulus; LTM, long-term memory; Phon/ortho parsing, phonological or orthographic parsing; X's com. intention, X's communicative intention. See text and Van Berkum (2018, 2019) for details.

is, and this can be evocative in a relatively unspecific way. Finally, the basic signs that make up a more complex utterance can carry their own emotional "payload" (Van Berkum, 2018). For example, if particular words such as "idiot," "bitch" or "disappointment" have often been used to implement negative social moves, to express negative stances, or to describe very negative situations, or if those words have often been witnessed to elicit strong negative responses in others, simple emotional conditioning will lay down negative affective connotations as part of the memory trace for that particular word, sometimes even to the extent that the word becomes taboo (Jay, 2009). As with other emotionally conditioned stimuli (e.g., a light that sufficiently reliably predicts a shock), the mere use of the word can then evoke a rapid automatic affective response, independent of the precise sentence-level message and the social intention of the speaker on this particular occasion.

If you hear an insult directed at somebody else, the ALC model leads us to expect that things will be partly different, but also partly comparable. If the person evaluated is somebody

you care for, the referential intention and associated social move may cause you to feel hurt or angry too, but if he or she is somebody you dislike, it might actually in part evoke a very different emotion (Singer et al., 2006). Other things might be less dependent on who the person evaluated is. For a species heavily invested in cooperation, for example, displays of an aggressive stance (such as a verbal or physical slap in the face) may automatically trigger negative emotion in the target of that aggression, as well as in those who witness somebody else being a target of aggression. Furthermore, and related, due to emotional conditioning, offensive words like "idiot," or "bitch" may elicit the same automatic affective response independent of who is being targeted, and what the word is used for in this specific communicative exchange.

## The Experiment

The goal of our study is to use EEG to examine whether verbal insults are less sensitive to repetition than compliments are, and, if so, which stage(s) of processing are implicated in the

adaptation, and which are not. In line with longstanding practices in EEG research, different EEG studies have used different latency ranges (and scalp regions) to quantify specific ERP components, possibly in part driven by the data. Recent work (Luck and Gaspelin, 2017) has drawn attention to the risks that are inherent to this flexibility in EEG analysis procedures. We therefore committed ourselves to only testing hypotheses in three predesignated critical latency ranges (and scalp regions) where, in our earlier EEG work (Van Berkum et al., 2009), we observed ERP effects of morally objectionable language: an "early" 200–250 ms latency range aimed at detecting P2 effects over the anterior part of the scalp, a "middle" 350–450 ms latency range aimed at detecting N400 effects over the central-posterior part of the scalp, and a "late" 500–650 ms latency range aimed at detecting LPP effects over the same central-posterior scalp region.

In line with the extant ERP literature, we made the following assumptions. First, insult-induced modulations of the word-elicited N400 component are assumed to index the impact of emotion on context-dependent lexical retrieval (or processes closely associated with that retrieval, see, e.g., Kutas and Federmeier, 2000, 2011; Lau et al., 2008; Van Berkum, 2009, 2012; Brouwer et al., 2017), involving usually left-dominant temporal as well as inferior frontal cortical generators (Halgren et al., 2002; Lau et al., 2008; Swaab et al., 2012). Second, an insult-induced increase of the preceding P2 will be taken as suggesting that emotional attention is up-regulating some aspect of visual stimulus processing (Delplanque et al., 2004; Huang and Luo, 2006; Hajcak et al., 2012; Carretié, 2014). Finally, an insult-induced increase of the LPP will be assumed to reflect additional—and possibly lengthy—elaborative processing of motivationally important, "significant" stimuli (e.g., Cacioppo et al., 1993, 1994, 2004; Schupp et al., 2000, 2004; Smith et al., 2003; Kisley et al., 2007; Sabatinelli et al., 2007, 2013; Holt et al., 2009; Frank and Sabatinelli, 2019; Hajcak and Foti, 2020), involving the extrastriate visual cortex as well as emotion-dedicated cortical areas (Frank and Sabatinelli, 2019), and possibly involving the activity of approach and avoidance systems (Bradley, 2009; Hajcak and Foti, 2020). Although it is tempting to assume that a "later" ERP effect like the LPP indexes later psycholinguistic computations as delineated in the ALC model (Van Berkum, 2019), early processing, such as the lexical retrieval of a word with strong emotional valence, may itself trigger additional processing in response to a motivationally relevant stimulus (i.e., a later LPP effect), regardless of whatever else happens in the psycholinguistic processing cascade laid out in the model.

In the experiment, participants read a series of repeated statements that realized three different speech acts: (1) insults, interpersonally evaluative statements that combined a person's name with a negative evaluative predicate (e.g., "Linda is horrible," "Paula is a liar"), (2) compliments, interpersonally evaluative statements that combined a person's name with a positive evaluative predicate (e.g., "Linda is impressive," or "Paula is an angel"), and (3) non-evaluative, factually correct descriptive statements (e.g., "Linda is Dutch," or "Paula is a student"). To examine whether the impact of the speech act depended on who the statement was about, half of these three sets of statements used the participant's own name, and the other half used the

name of somebody else. Although the study did not involve real interaction, we presented the statements as "being uttered by three specific men, about the participant herself, or some other participant in the study."

Our main goal was to look for signs of differential adaptation to repeated insults and repeated compliments, and to explore which stages of language-induced processing are implicated in any differential adaptation, and which are not. To examine this, we exposed our participants to 90 insults featuring their name as well as 90 insults featuring another person's name, and compared the ERPs elicited by those two types of insults (in the abovementioned P2, N400, and LPP latency ranges) to the ERPs elicited by 90 compliments featuring their name and 90 compliments featuring the other person's name, all as a function of repetition. Apart from the fact that the act of insulting was repeated many times, each particular statement item, e.g., "Linda is horrible," or "Paula is an angel," was presented many times throughout the session as well. Statements were distributed over several blocks, and we used the development of ERP amplitudes across these blocks as a measure of how participants adapted to repetition. Furthermore, we included 90 factually correct descriptive statements with the participant's name, and 90 more with the other person's name, to diversify the materials, and to also have an indication of how the response to non-evaluative and hence presumably "neutral" statements developed over blocks.

Apart from capturing attention, negative as well as positive emotional stimuli can also lead to arousal, that is, increased sympathetic activity within the autonomic nervous system that is intended to prepare the body for action as well as possible damage (Lang et al., 1993; Codispoti et al., 2001; Dawson et al., 2007). Arousal can be measured via the phasic skin conductance (SC) response, a measure that picks up on short-lived increases in sweat production (Lang et al., 1993; Codispoti et al., 2001; Dawson et al., 2007; Boucsein, 2012), and that has been shown to detect increases in arousal to reprimands and taboo words (Harris et al., 2003; Harris, 2004; Eilola and Havelka, 2011). In our study, skin conductance recording allowed us to assess the arousing potential of insults and compliments, again as a function of who was involved, and as a function of massive repetition.

Our predictions were as follows.

1. Based on our earlier EEG work with morally offensive language (Van Berkum et al., 2009), we expected verbal insults to generate a larger P2, a larger N400, and a larger LPP, all relative to compliments, but presumably also relative to neutral statements.
2. The massive repetition of evaluative statements should influence at least some of these three ERP responses to insults less than the corresponding ERP responses to compliments. So, for example, even though the amplitude of the insult-induced P2 might decrease over repetitions, the amplitude of the compliment-induced P2 should decrease over repetitions more, due to a stronger degree of adaptation to compliments. We had no strong predictions as to whether this attenuated adaptation for insults (relative to compliments) would involve the P2, the N400, or the LPP.

3. The differential impact of insults (relative to compliments, as well as neutral statements) should be larger for speech acts that described the participant than for speech acts that described somebody else, again in the P2, the N400, and/or the LPP.

4. We considered it likely that the differential adaptation to repetition for insults and compliments would depend on who the statement was about. In particular, we expected participants to adapt least to insults about themselves, and made no predictions about the other three cases.

5. With respect to skin conductance, the threat to the self that is posed by insults featuring the participant's own name led us to expect the strongest arousal (i.e., the highest phasic SC response) here, and we also expected these responses to be the most resistant to repetition. We had no strong predictions for the SC response to the other speech act types. Strongly positive stimuli are known to generate arousal too (e.g., Lang et al., 1993), so compliments to participants might lead to more arousal than compliments to somebody else, and neutral factual statements. However, if participants feel threatened by compliments to others, for example, a different arousal profile might arise.

## MATERIALS AND METHODS

### Participants

We tested 79 right-handed Dutch participants, all students at Utrecht University who identified themselves as female[2] ($M_{age}$ = 20.90, $Range_{age}$ = 18–30). They were non-dyslexic and did not report any neurological disorders. Participants who had a partner named "Bram," "Daan" or "Paul" could not sign up for the experiment, since those were the names of the men that supposedly uttered the statements. All research was conducted in line with the WMA Declaration of Helsinki—Ethical Principles for Medical Research Involving Human Subjects, as well as The Netherlands Code of Conduct for Scientific Practice issued in 2004 (revised in 2012) by the Association of Universities in the Netherlands (VSNU). The study was approved by the local Ethics Assessment Committee Linguistics (case number and relevant documents available upon request). Before the experiment, participants read a detailed information letter and accompanying informed consent form. These carefully explained the procedure and the offensive nature of some of the stimuli. We also made explicit, both in the consent form and face-to-face at the start of each session, that participants could decide to terminate the experiment at any time without having to provide a reason, and still receive the full financial compensation (10,- euro/hour). Before the experiment started the participant signed the informed consent form. None of the participants resigned.

### Materials and Design

To realize our Speech Act (insult, compliment, neutral description) by Person Described (participant, other) design, we created 10 unique insult frames (e.g., "<X> is horrible," "<X> is a liar"), 10 unique compliment frames (e.g., "<X> is impressive," or "<X> is an angel"), and 10 unique non-evaluative, factually

correct descriptive statements (e.g., "<X> is Dutch," or "<X> is a student"), and combined each with the participant's name or somebody else's name just before the experiment. Critical words, always in sentence-final position, were selected on the basis of a written pretest with 42 female Dutch students from Utrecht University ($M_{age}$ = 21.07, $Range_{age}$ = 18–28) who did not participate in the main experiment. Pretest participants read a large set of insulting, complimenting and non-evaluative neutral words embedded in a larger "<female-name> is <evaluative-predicate>" phrase, and rated the phrases on a seven-point scale ranging from "very insulting" (−3) to "neutral" (0) to "very complimenting" (3). The ratings were used to choose 10 strongly insulting words [e.g., "horrible," "liar," $M_{insults}$ = −2.39 (± 0.23)] as well as 10 equally strongly complimenting words [e.g., "impressive," "angel," $M_{compliments}$ = 2.45 (±0.18)], and 10 neutral words [e.g., "Dutch," "student," $M_{neutrals}$ = 0.05 (±0.10)] with strength expressed as deviation from the scale's neutral midpoint. The absolute deviation from the scale's neutral midpoint did not differ for insulting and complimenting word sets [$t_{(18)}$ = 0.64, $p$ = 0.53]. The word sets were also matched on length [$M_{insults}$ = 9.00 (±4.35), $M_{compliments}$ = 9.00 (±2.16), $M_{neutrals}$ = 9.00 (±3.16); $F_{(2,27)}$ = 0.00, $p$ = 1.00] and frequency [$M_{insults}$ = 49.45 (±78.91), $M_{compliments}$ = 55.28 (±56.08), $M_{neutrals}$[3] = 125.95 (±255.30); $F_{(2,27)}$ = 0.73, $p$ = 0.49] according to the SUBTLEX-frequency database for Dutch (Keuleers et al., 2010). In each set, half of the trials contained an indefinite article ("een" in Dutch), the other half did not. The full set of stimuli can be found in **Supplementary Table S1**.

Every trial started with a centered fixation-cross and a black screen paired with a beep, followed by a word-by-word presentation of the sentence. Each word was presented in the center of the screen in Calibri font size 35, with a duration that varied with word length (imitating natural reading parameters, see Otten and Van Berkum, 2007), and with the critical sentence-final (insulting, complimenting, or neutral) word shown for 1,000 ms in Calibri font size 50. After the critical word, another black screen was presented.

Statements were presented in blocks, with separate blocks for each of the Speech Act by Person Described combinations: insults involving the participant, compliments involving the participant, neutral observations involving the participant, insults involving somebody else, compliments involving somebody else, and neutral observations involving somebody else. In each block, each of the 10 relevant statements (e.g., 10 specific insults about the participant, such as "Linda is horrible") was presented three times, resulting in 30 stimuli per block. Within each block the 30 stimuli were arranged in pseudorandom order such that a specific insult was never presented twice in a row. Each of these six 30-item blocks was presented three times in a different pseudorandomized order, resulting in 18 blocks per session, and nine occurrences of any critical statement (e.g., "Linda is horrible," or "Paula is impressive") across the entire session. The order of the 18 blocks was counterbalanced according to a Latin

---

[2]We limited ourselves to a single gender to facilitate the construction of materials.

[3]Due to experimenter error, the frequency of one of the neutral words ("woman") was incorrectly coded, leading to a higher mean frequency of the neutral words set. Relative to the standard deviation across items, however, the mean difference was fairly small (as indicated by a nonsignificant test).

square procedure, resulting in six comparable stimulus lists. Half of these started with a statement block involving the participant's name, the other half with a statement block involving the length-matched name of somebody else. Orthogonal to this, one-third started with insults, one-third with compliments, and one-third with a neutral block. In all, each session contained 18 blocks of 30 statements each, which together lasted ~70 min. The effect of statement repetition was assessed *via* this Block (1–18) factor.

As they worked through each statement block, participants had no other task than to read each statement attentively. To keep participants sufficiently engaged, we also presented a very short auditory oddball task after each statement block, in which they had to detect 20 high beep targets (2,000 Hz) among 80 low beep distractors (1,000 Hz) and press a button when a target was detected as fast as possible. Each oddball minitask took 50 s.

## Procedure

After asking the participant for informed consent and explaining face-to-face that they could resign at any time without losing their compensation, the EEG and skin conductance (SC) electrodes were applied. Statement stimuli were presented using Neurobehavioral Systems' "Presentation" software, on a 22″ Iiyaman Prolite screen with a display resolution of 1,920 × 1,080 pixels and a 60 Hz refresh rate. Text was presented in Calibri font size 35, except for the critical word which was presented in font size 50, color 180, 180, 180 RGB. The distance between the screen and the participant was ~70 cm. At the start of the experiment the participant was introduced to four fictional characters: "In this experiment there are three men who will give some remarks, their names are Bram, Daan and Paul. Read their remarks carefully. The remarks will be about you or another woman. The other woman is a student in Utrecht and a participant in this experiment as well." This explanation was followed by six practice trials, which used evaluative words that were not in the experimental stimulus set. After the statement practice trials, the oddball task was explained and briefly practiced as well.

In the main session, evaluative or neutral factual statement blocks alternated with auditory oddball blocks. Prior to each statement block we mentioned which man the statements would come from in this block: always Bram for the neutral statements, and always Daan and Paul for the insults and compliments. We counterbalanced the latter across sessions such that a single man was presented as the source of insults to the participant and compliments to the other woman, and the other as the source of compliments to the participant and insults to the other woman (so that these men could be experienced as discriminative, rather than as, e.g., "always insulting everybody"). Blocks of statements and oddball were separated by a short break, and after six blocks of each task there was an obligatory longer break during which the experimenter checked how the participant was doing. Finally, the participant was asked to fill out an exit questionnaire and several personality questionnaires [the State-Trait Anxiety Inventory (STAI, Spielberger et al., 1983), the Rosenberg Self-Esteem Scale (RSES, Rosenberg, 1965), the Interpersonal Reactivity Index (IRI, Davis, 1983), the 10-item Connor-Davidson Resilience Scale (CD-RISC 10, Campbell-Sills and Stein, 2007), and the dimensions Outlook and Resilience

[Davidson and Begley, 2013)]. The questionnaires were added to investigate potential individual differences in an exploratory fashion. The entire experiment lasted ~2 h.

## EEG and Skin Conductance Measurements

EEG was recorded using 64 Ag/AgCl electrode BioSemi caps with a 10–20 configuration. We used two additional electrodes at the left and right mastoid for re-referencing during the analysis, two facial electrodes at the left and right outer canthi to measure the horizontal electrooculogram (hEOG), and another two electrodes above and below the left eye to measure vertical EOG (vEOG), in order to detect blinks and eye movements. Skin conductance was measured with two stainless-steel Nihon Kohden electrodes at the distal phalanx of the index finger and middle finger of the left hand. A BioSemi ActiveTwo system was used to sample and record all EEG and SC data at 2,048 Hz.

## EEG Analysis

The EEG data was preprocessed and analyzed offline using BrainVision Analyzer 2.1 software (Brain Products), Matlab (R2019a) and Rstudio (1.2.5042, R version 3.6.3). First, EEG signals were re-referenced to the average of the left and right mastoid, bandpass-filtered at 0.1–35 Hz (24 dB slope), and downsampled to 500 Hz. Next, all markers were checked using a customized macro, and segment-relevant condition markers were inserted at the onset of the critical word (CW). Next, the data was segmented into epochs from 200 ms before to 1,000 ms after CW onset, and segments were baseline-corrected by subtracting the mean signal amplitude in the CW-preceding 200 ms from all individual amplitude values in the remaining 1,000 ms. To optimize the signal-to-noise ratio, subsequent artifact rejection was done in two steps that focused on EOG and data channels, respectively. First, to remove eye movements and blinks, all epochs in which the bipolar hEOG and/or vEOG signal exceeded ±75 μV, or which displayed a voltage step of 50 μV or more between two neighboring sampling points, or in which the difference in signal activity was lower than 0.5 μV in an interval of 100 ms, were marked as bad trials. Second, the 64 regular EEG-channels were tested with the same criteria, but now only the specific channels failing the test were marked to be excluded from further analysis. As a final step of the preprocessing of the data the segments were exported to MAT-files for further processing in Matlab.

In Matlab all segments and channels within segments marked for exclusion were removed. Based on prior EEG work with verbal moral transgressions in our lab (Van Berkum et al., 2009), we computed mean amplitudes per segment and electrode for three predesignated latency-of-interest ranges that had also been used to assess P2, N400 and LPP effects in that earlier study: an "early" 200–250 ms latency range (optimized for detecting P2 effects), a "middle" 350–450 ms latency range (optimized for detecting N400 effects), and a "late" 500–650 ms latency range (optimized for detecting LPP effects)[4]. Furthermore, and

---

[4]As a compromise between strict replication and compatibility with the wider literature, we widened the rather narrow N400 latency range used in Van Berkum et al. (2009) with an extra 50 ms, while keeping it centered on the 400 ms mark (so from 375–425 to 350–450 ms).

again based on our earlier verbal moral transgressions study, we optimized our sensitivity for detecting any P2, N400 or LPP effects by computing, per segment, an average EEG amplitude over the anterior part of the brain (involving electrodes Fp1, AF3, AF7, F1, F3, F5, F7, FC1, FC3, FC5, FT7, Fp2, AF4, AF8, F2, F4, F6, F8, FC2, FC4, FC6, and FT8) for the 200–250 ms latency range, and an average EEG amplitude over the central and posterior part of the brain (involving electrodes C1, C3, C5, T7, CP1, CP3, CP5, TP7, P1, P3, P5, P7, PO3, PO7, O1, C2, C4, C6, T8, CP2, CP4, CP6, TP8, P2, P4, P6, P8, PO4, PO8, O2) for the 350–450 ms latency range as well as for the 500-650 ms latency range. These predesignated latency- and region-of-interest data operations condensed each of the participant's EEG signal in a critical epoch to just three hypothesis-relevant amplitude data points (anterior 200–250 ms, posterior 350–450 ms, and posterior 500–650 ms), all exported into a text file for further analysis using Rstudio.

The three latency ranges were analyzed separately in Rstudio using linear mixed models. To further reduce noise, we first averaged the three repetitions of each specific critical item (e.g., "Linda is horrible") within each block, such that three repetition averages remained (one for each relevant block). The participant and statement variances were included as random factors and were estimated simultaneously, resulting in a cross-classified model (Quené and van den Bergh, 2004, 2008). For each latency range, we constructed models for the mean amplitude by iteratively adding potentially relevant components and testing for significant model improvement at each addition (using the likelihood ratio (−2LL difference chi-square) test, $p < 0.05$). We started by comparing an empty model, containing only the random factors for participant and statement, to the base model containing the random factors and the three main effects of interest: Speech Act (insult, neutral, compliment), Person Described (you, other) and Block (1–18). The factor Block was centered to avoid correlation between the linear and quadratic effect that was added in the next step. We then added another factor in each next step, starting with the 2-way interactions and working all the way up to the 4-way interaction. The final model was in fact the full factorial model. We then evaluated which model fit the data best and only kept those factors that explained a significant amount of variance once they were added to the model or which were necessary to test hypothesized interactions (Winter, 2019). Factors that did not significantly improve the model were dropped (H. van den Bergh, personal communication; Seasholtz and Kowalski, 1993). The summary of arriving at the best model is given in **Table 1**. The best and most sparse model was used to estimate the model parameters correctly. For transparency, the full factorial models are also reported in **Supplementary Tables S2–S4**. For the *post-hoc* analyses we used Tukey HSD corrected pairwise comparisons.

Figures with results from the linear mixed model analysis, reporting the mean amplitudes for each latency range, were based on parameter estimates from the best models, generated using Rstudio. The ERP and topography results were generated in BrainVision Analyser. For visualization the data were processed from the segmentation step as follows. The data were segmented

into the 18 blocks of the experiment. For each block the data was then segmented into epochs from 200 ms before to 1,000 ms after the onset of every critical word, and segments were baseline-corrected by subtracting the mean signal amplitude in the first 200 ms of each epoch from all individual amplitude values in that epoch. Next, the artifact rejection procedure based on the EOG-channels was performed with similar settings as mentioned above. However, instead of only marking the bad trials, they were removed from the analysis right away. This was also done for the artifact rejection based on the remaining 64 electrodes. The channels that were marked as bad were immediately excluded from further analysis. As a final step, the segments were averaged per block. For visualization purposes new aggregate channels representing four quadrants were calculated by averaging the following channels. A Left Anterior (LA) channel was calculated from electrodes Fp1, AF3, AF7, F1, F3, F5, F7, FC1, FC3, FC5, and FT7; a Left central/Posterior (LP) channel was calculated from electrodes C1, C3, C5, T7, CP1, CP3, CP5, TP7, P1, P3, P5, P7, PO3, PO7, and O1; a Right Anterior (RA) channel was calculated from electrodes Fp2, AF4, AF8, F2, F4, F6, F8, FC2, FC4, FC6, and FT8; and a Right Central/Posterior (RP) channel was calculated from electrodes C2, C4, C6, T8, CP2, CP4, CP6, TP8, P2, P4, P6, P8, PO4, PO8, and $O_2$. The average response per Speech Act by Person Described and by number of repetition (1–3) was computed for these four channels.

## SC Analysis

The SC data was preprocessed offline using BrainVision Analyzer 2.1 software (Brain Products). The data was downsampled to 500 Hz and segmented into epochs from the onset until 165 s after each Speech Act by Person Described block, this epoch included the entire block of 30 trials. These epochs were exported into mat-files and further analyzed using Matlab R2019b and the Ledalab toolbox (Benedek and Kaernbach, 2010). We used the Continuous Decomposition Analysis to decompose the phasic and tonic activity. The data was down sampled to 10 Hz and adaptive smoothing using a Gaussian window was applied. Six sets of initial values were used to optimize the solution. The event-related phasic activity was further analyzed. Since the final 30th trial in each block did not always have 4 s of clean signal after critical word onset, which was necessary for further analysis, we only analyzed the first 29 trials. We used a 10 μS threshold for significant phasic activity and computed the mean activity in the time window of 1–4 s after the critical event as the area under the curve (squared integrated skin conductance response: $ISCR^2$). The results were log transformed to normalize the data, and exported to a text-file for further analysis using Rstudio. The linear mixed model analysis in Rstudio followed the same procedure as reported above for the EEG analysis: the models were built iteratively until the full-factorial model was reached and subsequently the most sparse model was achieved by removing non-significant factors that did not improve the model fit. This procedure resulted in the best, most sparse model (see **Supplementary Table S5**).

**TABLE 1 |** Linear mixed models analysis summary.

| Predictors | 200–250 (M1b, $df$ = 13,896) | | | 350–450 (M4a, $df$ = 13,892) | | | 500–650 (M11g, $df$ = 13,891) | | | SCR (M9d, $df$ = 14,029) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Estimates (95% CI) | z | p | Estimates (95% CI) | z | p | Estimates (95% CI) | z | p | Estimates | z | p |
| (Intercept) [compliment] | 6.55 (5.89, 7.20) | 19.58 | **<0.001** | 2.62 (2.03, 3.21) | 8.75 | **<0.001** | 2.84 (2.28, 3.41) | 9.79 | **<0.001** | 0.22 (0.19, 0.25) | 13.74 | **<0.001** |
| SpeechAct [neutral] | 0.08 (−0.32, 0.47) | 0.38 | 0.703 | 0.24 (−0.21, 0.69) | 1.06 | 0.288 | 0.09 (−0.43, 0.61) | 0.33 | 0.742 | 0.01 (0.01, 0.02) | 3.97 | **<0.001** |
| SpeechAct [insult] | 0.76 (0.36, 1.16) | 3.69 | **<0.001** | 0.65 (0.19, 1.11) | 2.78 | **0.005** | 0.35 (−0.18, 0.88) | 1.29 | 0.198 | −0.00 (−0.01, 0.00) | -0.54 | 0.591 |
| Person [3rd person] | | | | −0.54 (−0.91, −0.17) | -2.83 | **0.005** | −0.23 (−0.74, 0.27) | -0.90 | 0.368 | 0.01 (−0.00, 0.01) | 1.51 | 0.131 |
| BlockCent | −0.81 (−1.01, −0.60) | -7.83 | **<0.001** | −0.52 (−0.87, −0.16) | -2.83 | **0.005** | −0.65 (−1.04, −0.26) | -3.27 | **0.001** | −0.02 (−0.04, −0.01) | -3.11 | **0.002** |
| Block2Cent | | | | 0.50 (0.05, 0.95) | 2.20 | **0.028** | 1.69 (1.00, 2.38) | 4.80 | **<0.001** | 0.10 (0.08, 0.12) | 12.55 | **<0.001** |
| SpeechAct [neutral] * BlockCent | | | | −0.52 (−1.03, −0.02) | -2.04 | **0.041** | −0.54 (−1.08, 0.01) | -1.93 | **0.054** | | | |
| SpeechAct [insult] * BlockCent | | | | −0.80 (−1.31, −0.28) | -3.04 | **0.002** | −1.04 (−1.60, −0.48) | -3.66 | **<0.001** | | | |
| Person [3rd person] * BlockCent | | | | | | | | | | 0.02 (0.01, 0.03) | 3.28 | **0.001** |
| Person [3rd person] * Block2Cent | | | | | | | −1.24 (−2.22, −0.27) | -2.50 | **0.013** | −0.06 (−0.09, −0.04) | -5.62 | **<0.001** |
| BlockCent * Block2Cent | | | | | | | | | | −0.08 (−0.10, −0.05) | -6.10 | **<0.001** |
| | **Random effects** | | | **Random effects** | | | **Random effects** | | | **Random effects** | | |
| $\sigma^2$ | 39.55 | | | 41.98 | | | 49.41 | | | 0.03 | | |
| $\tau_{00 \ Participant}$ | 7.20 | | | 3.96 | | | 2.15 | | | 0.02 | | |
| $\tau_{00 \ item}$ | 0.24 | | | 0.35 | | | 0.51 | | | | | |
| ICC | 0.16 | | | 0.09 | | | 0.05 | | | 0.43 | | |
| $N_{Participant}$ | 79 | | | 79 | | | 79 | | | 78 | | |
| $N_{item}$ | 60 | | | 60 | | | 60 | | | | | |
| Deviance | 90912.204 | | | 91702.916 | | | 93922.279 | | | −11058.841 | | |
| Log-Likelihood | −45456.102 | | | −45851.458 | | | −46961.140 | | | 5529.420 | | |

*Best models for the analysis of mean ERP amplitudes in the three ERP latency ranges of interest and for the analysis of mean skin conductance response (SCR). Significant p-values are given in bold.*

## Individual Differences Analysis

In an additional exploratory analysis, we added the standardized scores of our various personality trait measures [the State-Trait Anxiety Inventory, the Rosenberg Self-Esteem Scale, the Interpersonal Reactivity Index, the 10-item Connor-Davidson Resilience Scale, and two Outlook and Resilience dimensions proposed by Davidson and Begley (2013)], to each best model (derived as described above) and tested whether the extended model had a better fit to the data. None of the personality trait measures significantly improved the model fit, neither in the analysis of any of the three ERP time windows nor in that of skin conductance (all $p$'s > 0.45, see our online repository for further details). Because these are null results in exploratory analyses, we will refrain from deriving any strong inferences from them.

## RESULTS

We first discuss all effects of Speech Act, for which we present mean EEG amplitude in the P2, the N400, and the LPP latency ranges, followed by the phasic skin conductance response (SCR). After this, we do the same for all Speech Act by Block effects, and then we discuss all interaction effects involving Person Described.

## Speech Act

**Figure 2** displays grand average ERPs, insult-compliment scalp topographies, and mean SC responses, the latter two both for the entire session as well as for the three relevant blocks (e.g., the three blocks with insults involving the participant).

Based on our earlier EEG work with morally offensive language (Van Berkum et al., 2009), we expected verbal insults to generate a larger P2, a larger N400, and a larger LPP, all relative to compliments, but presumably also relative to neutral statements. This factor was therefore included in all best final models, see **Table 1**.

### P2 (200–250 ms, Anterior Region)

In line with our predictions, mean ERP amplitude in the P2 latency range was more positive to insults than to compliments ($b_{ins-com} = 0.76, z = 3.69, p_{Tukey} < 0.001$), and also more positive to insults than to neutral descriptions ($b_{ins-neu} = 0.68, z = 3.35, p_{Tukey} = 0.002$). Mean ERP amplitude in the P2 latency range for compliments and neutral descriptions did not differ ($b_{com-neu} = -0.08, z = -0.38, p_{Tukey} = 0.92$). As can be seen in **Figure 2A**, the differential insult effect, relative to compliments, has a symmetrical scalp distribution with a fronto-central maximum, consistent with other reports of emotion-induced P2 effects. We therefore interpret this as a true P2 effect.

### N400 (350–450 ms, Central and Posterior Region)

Against our expectations, insults elicited a less negative mean ERP amplitude in the N400 latency range than compliments did ($b_{ins-com} = 0.65, z = 2.79, p_{Tukey} = 0.01$). Mean ERP amplitude in the N400 latency range for insults and neutral descriptions did not differ ($b_{ins-neu} = 0.41, z = 1.76, p_{Tukey} = 0.18$), nor did it for compliments vs. neutral descriptions ($b_{com-neu} = -0.24, z = -1.07, p_{Tukey} = 0.53$).

### LPP (500–650 ms, Central and Posterior Region)

Also against our expectations, mean ERP amplitudes in the LPP latency range did not differ between insults and compliments ($b_{ins-com} = 0.35, z = 1.30, p_{Tukey} = 0.40$). Mean ERP amplitude in the LPP latency range for insults and neutral descriptions also did not differ ($b_{ins-neu} = 0.26, z = 0.98, p_{Tukey} = 0.59$), nor did it for compliments vs. neutral descriptions ($b_{com-neu} = -0.09, z = -0.33, p_{Tukey} = 0.94$).

## Skin Conductance

Finally, mean SC response amplitudes to insults and compliments also did not differ ($b_{ins-com} = -0.00, z = -0.54, p_{Tukey} = 0.85$). Surprisingly, both of these evaluative speech acts led to a lower SCR amplitude than neutral, factually correct descriptions ($b_{ins-neu} = -0.01, z = -4.46, p_{Tukey} < 0.001; b_{com-neu} = -0.01, z = -3.97, p_{Tukey} < 0.001$).

## Summary

In sum, across the entire session and pooled over who is described, we see insults elicit a larger P2 than compliments, but not a larger LPP, nor a larger SC response. Also, in the N400 latency range, the insult-elicited ERP-signal is more positive than the compliment-elicited ERP-signal. Note that the waveforms in **Figure 2** do suggest a more negative ERP-signal for insults than compliments around 300 ms, a timing that falls outside the scope of our predesignated latency ranges of interest. Thus, although descriptively the triphasic positive-negative-positive differential effect exhibited by the ERPs for insults vs. compliments resembles that for morally objectionable vs. acceptable statements in our earlier study (Van Berkum et al., 2009), statistical tests in the same latency ranges only confirm the early P2 effect.
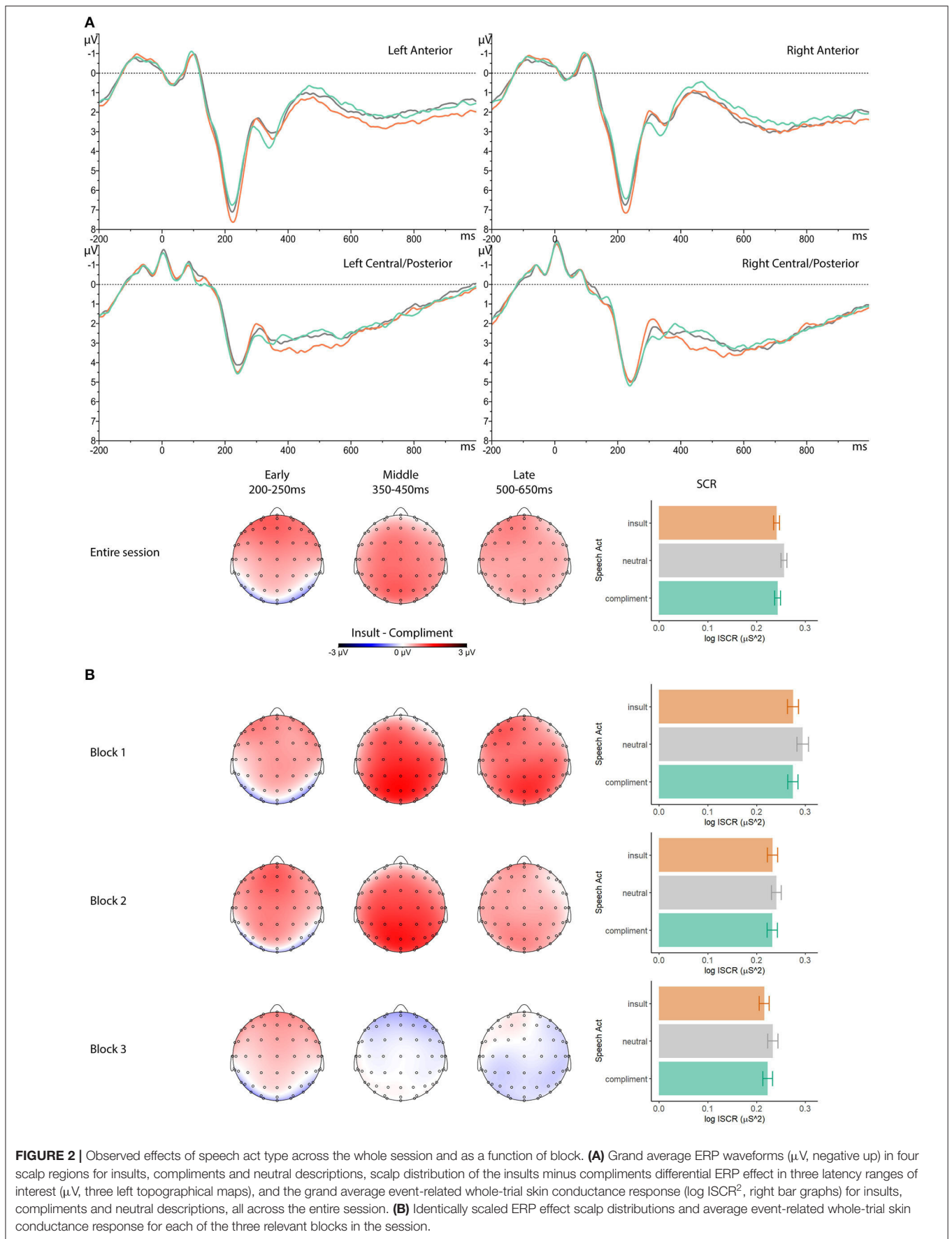
Next, we examine to what extent these results are qualified by repetition of the speech acts (Block), by who is implicated (Person Described), and by the interaction between these two factors.
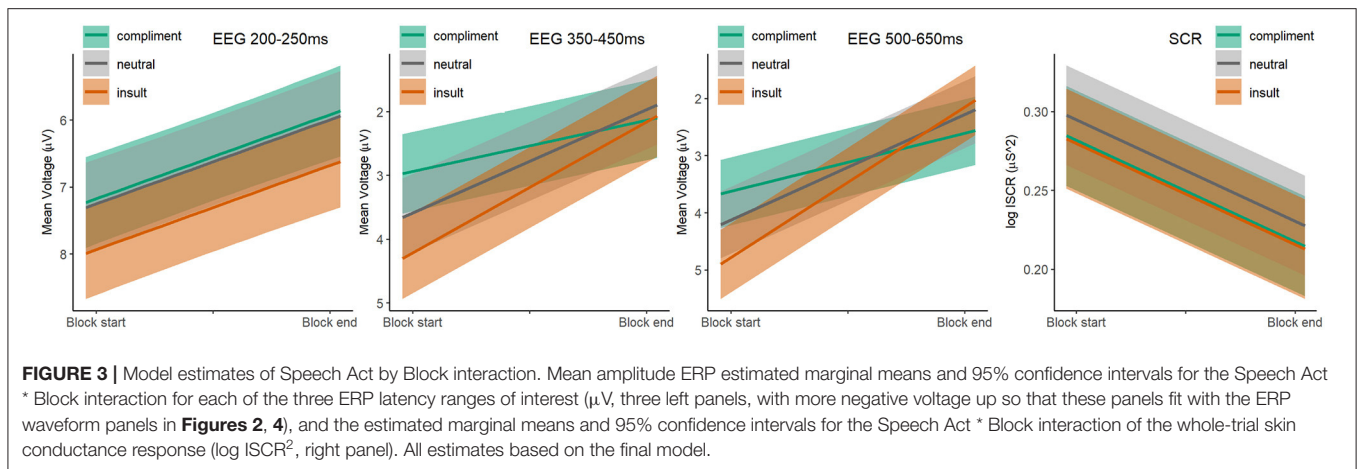
## Speech Act as a Function of Block

In line with our critical hypothesis of a differential adaptation to repeated verbal insults and repeated compliments, we predicted that the massive repetition of these evaluative speech acts would affect at least some of the three ERP responses to insults less than it would affect the corresponding ERP responses to compliments. **Figure 3** shows the relevant model estimates (as such complementing the observed data shown in **Figure 2B**).

### P2 (200–250 ms, Anterior Region)

As for mean ERP amplitude in the P2 latency range, the predicted differential adaptation to repetition was not observed. The best model included only the main effect of Speech Act and the main effect of Block, but no Speech Act by Block interaction, nor any interaction involving Person Described (see **Table 1**). As shown in **Figure 3**, mean ERP amplitudes to insults, compliments and neutral descriptions adapt over repetition in the same non-zero linear way, becoming increasingly less positive as the session progresses ($b = -0.81, z = -7.83, p_{Tukey} < 0.001$), with no significant differences between the trends. As can be seen in **Figure 2B**, and consistent with the latter, the size and scalp distribution of the differential P2-effect between insults and

**FIGURE 2 |** Observed effects of speech act type across the whole session and as a function of block. **(A)** Grand average ERP waveforms (μV, negative up) in four scalp regions for insults, compliments and neutral descriptions, scalp distribution of the insults minus compliments differential ERP effect in three latency ranges of interest (μV, three left topographical maps), and the grand average event-related whole-trial skin conductance response (log ISCR$^2$, right bar graphs) for insults, compliments and neutral descriptions, all across the entire session. **(B)** Identically scaled ERP effect scalp distributions and average event-related whole-trial skin conductance response for each of the three relevant blocks in the session.

**FIGURE 3 |** Model estimates of Speech Act by Block interaction. Mean amplitude ERP estimated marginal means and 95% confidence intervals for the Speech Act * Block interaction for each of the three ERP latency ranges of interest (μV, three left panels, with more negative voltage up so that these panels fit with the ERP waveform panels in **Figures 2**, **4**), and the estimated marginal means and 95% confidence intervals for the Speech Act * Block interaction of the whole-trial skin conductance response (log ISCR$^2$, right panel). All estimates based on the final model.

compliments does not change as a function of repetition, and is clearly as visible in the last block as it is in the first one.

### N400 (350–450 ms, Central and Posterior Region)

As for mean ERP amplitude in the N400 latency range, insults, compliments and neutral descriptions responded differently to repetition, albeit not in the way we had expected. The positive mean amplitudes in this latency range became less positive in a statistically linear way for all three speech act types, but with a faster rate of adaptation for insults ($b = -1.31$) than for compliments ($b = -0.52$, $b_{ins-com} = -0.80$, $z = -3.04$, $p_{Tukey} = 0.01$). As can be seen in **Figure 2B**, the widespread differential positivity for insults, compared to compliments, was visible in the first two blocks but completely absent in the last block. The rate of adaptation for insults did not statistically differ from that of neutral descriptions ($b = -1.04$; $b_{ins-neu} = -0.27$, $z = -1.06$, $p_{Tukey} = 0.54$), and the rate of adaptation for compliments also did not statistically differ from that of neutral descriptions ($b_{com-neu} = 0.52$, $z = 2.04$, $p_{Tukey} = 0.10$).

### LPP (500–650 ms, Central and Posterior Region)

As for mean ERP amplitude in the LPP latency range, insults, compliments and neutral descriptions again responded differently to repetition, in a way that actually echoes responses in the N400 latency range. As in that preceding latency range, the positive mean amplitudes in the LPP latency range became less positive in a statistically linear way for all three speech act types, but with a faster rate of adaptation for insults ($b = -1.69$) than for compliments ($b = -0.65$, $b_{ins-com} = -1.04$, $z = -3.66$, $p_{Tukey} < 0.001$). As can be seen in **Figure 2B**, and consistent with the relevant slopes in **Figure 3**, the widespread differential positivity for insults, compared to compliments, was visible in the first two blocks but completely absent in the last block. The rate of adaptation for insults did not statistically differ from that of neutral descriptions ($b = -1.18$; $b_{ins-neu} = -0.51$, $z = -1.79$, $p_{Tukey} = 0.17$), and the rate of adaptation for compliments also did not statistically differ from that of neutral descriptions ($b_{com-neu} = 0.54$, $z = 1.93$, $p_{Tukey} = 0.13$).
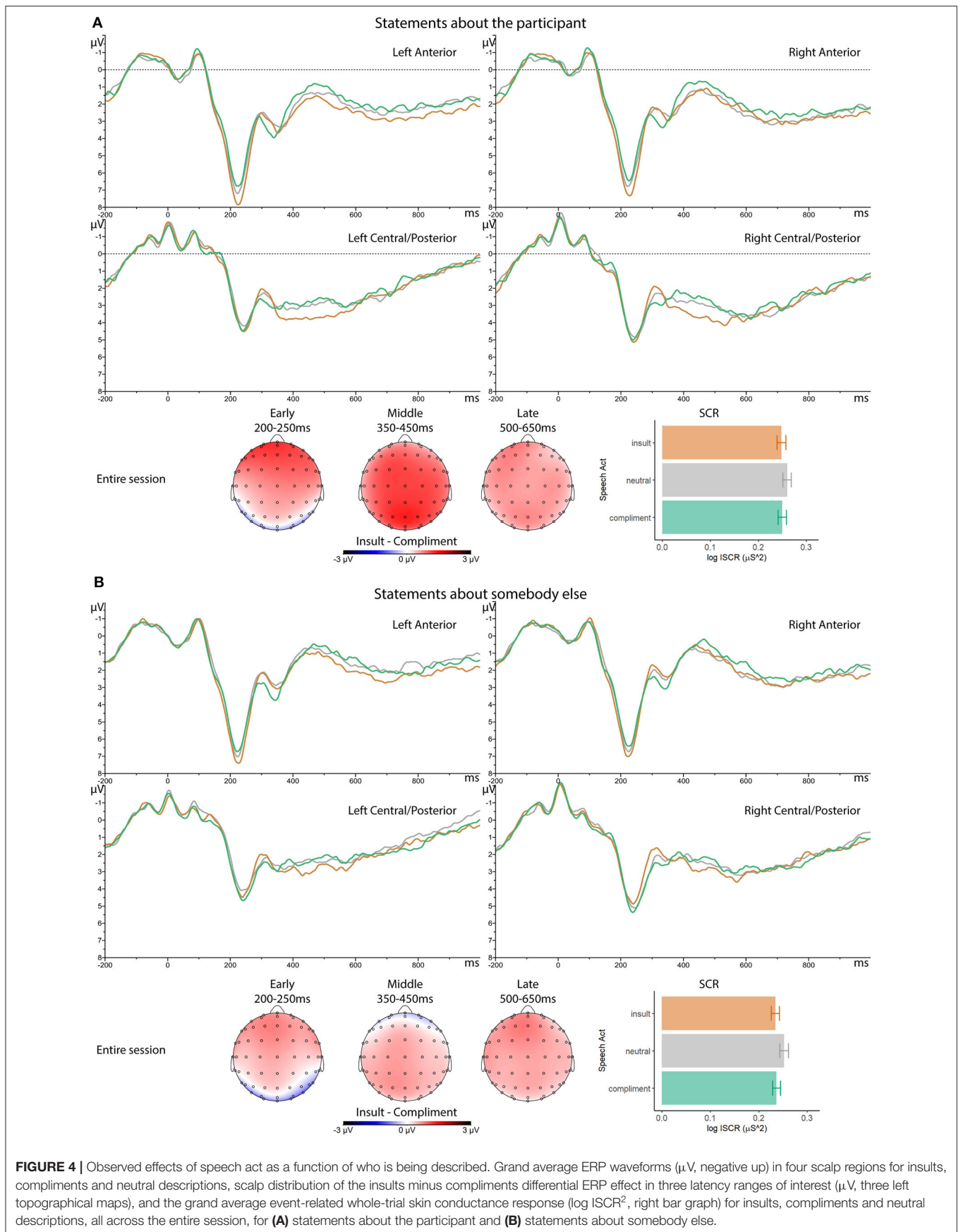
### Skin Conductance

Mean SC responses to insults, compliments and neutral descriptions all declined with repetition, at the same linear rate ($b = -0.03$, $z = -9.05$, $p_{Tukey} < 0.001$). As can be seen in **Figure 2B**, the unexpected skin conductance pattern that was observed across the entire session was also present in each of the three blocks, with the highest SC response for neutral descriptors, and no indications of differential arousal elicited by insults and compliments.

### Summary

In all, our results do not provide any evidence that the massive repetition of evaluative speech acts affects responses to insults less than it affects responses to compliments. Although the P2 adapts, it does so equally for both, leaving the differential effect intact. Also, although there is differential adaptation in the N400 and LPP latency ranges, insult-induced responses adapt more than compliment-induced responses, not less. The result is that the increased ERP positivity to insults relative to compliments observed in both latency ranges has fully disappeared by the last block. The associated ERP waveforms in **Figure 2A**, the comparable scalp distributions of the insult minus compliment effect in the 350–450 ms and 500–650 ms latency ranges displayed in **Figure 2A** and, per block, in **Figure 2B**, as well the virtually identical adaptation slopes in **Figure 3** all suggest that the most parsimonious account is one that involves the same long-lasting insult-induced positive shift, already emerging in the 350–450 ms latency range, continuing in slightly attenuated form in the 500–650 ms latency range, and vanishing for both ranges in the last block. Finally, skin conductance responses to insults, compliments, and neutral descriptors declined over blocks in a similar linear way, while retaining the unexpected elevation of arousal to neutral descriptions.

### Interactions With Person Described

As revealed by the fact that none of the best statistical models retained predictors involving Person Described (see **Table 1**), the picture painted above is not significantly modulated by who the insults, compliments or neutral descriptions are about. This means that the repetition-robust P2 effect to insults relative to compliments did not depend on who was involved, and

**FIGURE 4 |** Observed effects of speech act as a function of who is being described. Grand average ERP waveforms ($\mu$V, negative up) in four scalp regions for insults, compliments and neutral descriptions, scalp distribution of the insults minus compliments differential ERP effect in three latency ranges of interest ($\mu$V, three left topographical maps), and the grand average event-related whole-trial skin conductance response (log ISCR[2], right bar graph) for insults, compliments and neutral descriptions, all across the entire session, for **(A)** statements about the participant and **(B)** statements about somebody else.

nor did the repetition-sensitive differences between insults and compliments in later (350–450 and 500–650 ms) latency ranges. Even the unexpected higher arousal to neutral descriptors, relative to evaluative ones, did not change when statements involved the participant, rather than somebody else. The equivalence of ERP and SCR effects is illustrated in **Figure 4**. Note, in particular, the very similar triphasic positive-negative-positive response in the waveforms, and the similar P2 effect scalp distributions. Although the various insult effects seem somewhat weaker when the evaluative statements involve somebody else, this is not corroborated by the statistics.

## DISCUSSION

Genuine insults violate a universal moral imperative not to inflict harm on others (Greene, 2013). Insults also threaten face or reputation, which for members of an ultrasocial species should be a highly evocative event (Goffman, 1967; Leary, 2005). We used EEG and skin conductance to explore the short-term impact of verbal insults, compared to that of verbal compliments as well as neutral factual descriptions. Partly based on our earlier ERP work on other verbal violations of moral norms (Van Berkum et al., 2009), we predicted that insults would elicit a larger P2, a larger N400, and a larger LPP than compliments (and neutral factual statements), as well as more elevated physiological arousal (see, e.g., Harris, 2004; Eilola and Havelka, 2011). We also expected that at least some of these differential insult effects would be larger when statements involved the participant, instead of somebody else. Most critically, we predicted that the EEG and skin conductance response to insults would be less affected by repetition than the equivalent responses to other speech act types.

### The Early ERP Effect

The most robust insult effect in our study is the enhanced positivity emerging between 200 and 250 ms after onset of the critical insulting word. In light of prior ERP research, the most parsimonious account of this early effect is that it is a modulation of the P2 component, an anteriorly maximal short-lived ERP component that is sensitive to manipulations of attention (Luck and Hillyard, 1994; Hajcak et al., 2012). The exact functional significance and neural origin of the P2 is not yet well-understood, but research suggests that it originates early in the visual processing stream, and that an increase in amplitude indexes additional attention that is rapidly, and automatically, captured by emotional or unexpected events (Delplanque et al., 2004; Huang and Luo, 2006; Hajcak et al., 2012; Carretié, 2014). Given that insults are moral transgressions as well as verbal slaps in the face, it makes sense to find that they capture attention rapidly, at least to a larger degree than verbal compliments and neutral descriptions do.

This differential P2 effect does not change in the face of massive repetition (see **Figures 2B**, **3**). Even after two-thirds of the experiment during which readers had already seen a respectable 120 verbal insults (during which each of 10 unique insulting critical words had already been used 12 times), insults continued to generate a differential P2 effect of the same magnitude, and with the same scalp distribution. This suggests

that whatever aspect of insults is responsible for capturing extra attention does so in a highly robust way. Although the stability of this differential P2 effect testifies to the fact that insults capture additional attention in a robust, repetition-insensitive way (at least within the bounds of repetition examined in this experiment), it also disconfirms our prediction that compliments would adapt sooner than insults, as the latter would predict an increasing differential effect.

Interestingly, our P2 findings suggest that the extra attention capture indexed by this effect occurs to an equal extent when insults involved the participant or somebody else (see **Figure 4**). Also in contrast to our expectations, there is no evidence that at the level of the P2, insults involving the participant adapt less to repetition than insults involving somebody else. The insensitivity of the P2 to who is being evaluated in insults and compliments suggests that the processes reflected in this early component are not influenced by the output of the compositional and referential processes that are required to compute the difference between, say, "you are a liar" and "she is a liar."

In terms of a generally accepted cascade of psycholinguistic processes (as captured, e.g., in the ALC model; Van Berkum, 2018, 2019), the most obvious explanation is that the insult-induced P2 effect has a lexical source, and simply reveals the brain's increased sensitivity to negative evaluative words, such as "liar," "disappointment," or "bitch," as the affective meaning of these signs is retrieved from long-term memory. In our study, the words used in the insults were really very negative evaluative words ("ugly," "horrible" "disappointment"), and some of them even involved taboo words (e.g., "bitch," "whore"). Such words have been observed to rapidly deliver part of that strong negative "payload"—such as by immediately capturing attention and recruiting additional processing—as part of lexical retrieval, even when presented in isolation (e.g., Wabnitz et al., 2012, 2016; see Citron, 2012, for a review). Of course, very positive words will rapidly deliver their affective meaning as part of lexical retrieval too (Citron, 2012). In our study, however, the positive evaluative words elicited a significantly smaller P2 than the negative ones, and the P2 to positive words did not differ from that elicited by critical words in factual descriptions. The P2 results in our study therefore not only reveal a stable early differential insult effect, but also testify to a negativity bias in the amount of attention that is automatically allocated to very negative vs. very positive interpersonal evaluations. Whether this should be taken to reflect valence, arousal, or a combination thereof is something we return to later.

### Later ERP Effects

In earlier work (Van Berkum et al., 2009; see also Foucart et al., 2015; Hundrieser and Stahl, 2016; Peng et al., 2017), critical words in morally objectionable survey statements have elicited a larger N400 than their counterparts in morally acceptable statements. A study that compared phrasal insults to compliments (Otten et al., 2017) has revealed an N400 effect as well. In the current study, however, insulting words did not elicit a larger N400, and instead elicited a larger positivity between 350 and 450 ms. The ERP waveforms in **Figures 2** and **4** do reveal a somewhat earlier short-lived negative deflection for

insulting words relative to complimenting words, around 300 ms. However, neutral factual statements descriptively elicit the same negativity, and the difference vanished before the canonical 400 ms mark.

This unexpected early negativity does not resemble the occipito-temporal EPN often reported for emotionally competent words (see Citron, 2012; Frank and Sabatinelli, 2019). The timing of this unexpected negativity falls outside our predesignated latency ranges of interest, and to not inflate error rates, we refrain from *post-hoc* statistics to corroborate it. **Figure 4** does clearly reveal that, at a descriptive level, the effect emerged in all quadrants, for speech acts involving the participant as well as somebody else. This stability bodes well for future, more hypothesis-driven examinations.

Based on our findings for morally offensive language (Van Berkum et al., 2009), and in line with other studies revealing a larger LPP for negative over positive stimuli (e.g., Schupp et al., 2004; Carretié et al., 2008; Herbert et al., 2008; Holt et al., 2009; Kissler et al., 2009; Hajcak et al., 2013; Sabatinelli et al., 2013), we had also predicted insulting words to elicit a larger LPP than compliment words (and neutral factual words). We did not obtain this insult LPP effect in our predesignated latency range of 500–650 ms. However, the waveforms shown in **Figures 2A** and **4** suggest a single insult-induced central-posterior positivity that starts around 350 ms after critical word onset and lasts until at least 600 ms. This idea is corroborated by the fact that in the 350–450 ms and 500–650 ms latency ranges, the insult effects have the same scalp distribution, as well as by the fact that mean amplitudes in the conditions involved adapt to repetition in the same way. LPP effect onsets around 300 ms are not uncommon, and even earlier LPP effects have occasionally been reported (emerging as early as 160 ms, Hajcak and Foti, 2020). The most parsimonious interpretation, therefore, is that the significant insult-induced positivity that we observe in the 350–450 ms latency range in our study is the start of an early-onsetting LPP effect[5].

The presence of an insult-induced LPP effect resonates with other work in which insults were contrasted with compliments (Otten et al., 2017), and presumably reflects rapidly increased elaborative processing of motivationally important stimuli (e.g., Cacioppo et al., 1993, 1994, 2004; Schupp et al., 2000, 2004; Smith et al., 2003; Kisley et al., 2007; Sabatinelli et al., 2007, 2013; Holt et al., 2009; Frank and Sabatinelli, 2019). However, unlike the increased emotional attention reflected in the early P2 effect, the increased elaborative processing of insults reflected in the later LPP did not withstand repetition, and had disappeared by the end of the experiment. This goes against our expectation that repeated insults might keep their sting longer than repeated compliments, and in fact reveals that at the level of the LPP, if anything, insults lose their impact on processing more rapidly

than compliments. Also, this happens equally for statements involving the participant and for those involving somebody else, disconfirming our additional expectation that insults directed at the participant would adapt least of all. In fact, who was evaluated did not matter at all to LPP amplitudes in the 350–450 and 500–650 ms latency range, neither for trends over time and repetition, nor for the insult LPP effect pooled over the entire session. The hint of a stronger insult effect for insults directed at participants that is visible in **Figure 4** is not corroborated by the statistics.

## Skin Conductance Effects

The skin conductance findings are puzzling. We expected that the threat to the self that is posed by insults featuring the participant's own name would generate the strongest arousal (i.e., the highest skin conductance response). In our study, however, it is the factually correct neutral descriptions that generated the strongest arousal, while insults and compliments did not differ in their arousing potential. As illustrated in **Figures 2–4** and confirmed by the statistics, this factual description effect in arousal remained the same throughout the session, and also did not differ as a function of who was involved. One explanation for the unexpected arousal to neutral factual descriptions might be that, as non-evaluative statements, they occur twice as infrequently in our experiment as the evaluative statements, and might as such perhaps require additional processing (see Citron et al., 2013, for a comparable account). However, we presented quite a few neutral statements in a single session (180), and grouped our statements in blocks of 30 of the same type, two factors that would seem to go against a simple frequentist oddball account. Furthermore, one would have to explain why these neutral oddballs noticeably increase arousal but not, for example, the P2, or the ERP amplitude in subsequent latency ranges (Polich, 2012; Hajcak and Foti, 2020).

We suspect that things are more complex than a simple frequency-based oddball account would suggest. In the context of many interpersonally evaluative statements like "Linda is horrible," "Paula is a liar," "Linda is impressive," and "Paula is an angel," non-evaluative, factually correct descriptive statements like "Linda is a person" or "Paula is a woman" may perhaps take on a meaning that makes them more complex, and more ambiguous, than we intended, and might as such have led to more arousal. Furthermore, in retrospect, some of our neutral statements, such as the 18 statements describing someone as a "person," were slightly odd—when would one ever say such a thing? Of course, this account is entirely *post-hoc*, we had not predicted our factually correct statements to generate more arousal than insults and compliments. However, we note that psycholinguistics has a long history of rediscovering that intendedly neutral stimuli are not as simple as one hoped for (see, e.g., the neutral stimuli in Rohr and Abdel Rahman, 2018). Our study may be another case in point.

Perhaps more importantly, these skin conductance results do not pattern with those in any of the ERP latency ranges that we examined in our study. This is informative as to what the insult-induced P2 and LPP effects reflect. There is an interesting debate on whether the effects of emotional words involve valence, arousal, or both (see, e.g., Bayer et al., 2010; Citron et al., 2013;

---

[5]In their work on morally objectionable statements, Van Berkum et al. (2009) left open the possibility that the early transgression-induced ERP positivity around 200–250 ms might also be part of an early-onsetting LPP effect, momentarily canceled out by an N400 effect. However, because the two insult-induced positivities in our study adapt in a different way, and also differ in scalp distribution, this scenario can be reasonably ruled out here.

Recio et al., 2014). In the absence of a measurable difference in physiological arousal (in an otherwise sensitive SCR measure), it is tempting to infer that our insult-induced P2 and LPP effects reflect a strong difference in valence, or some other aspect that captures the higher significance (Bradley, 2009; Hajcak and Foti, 2020) of the insults that we used.

## The Bigger Picture

This brings us to an important question: what is the significance of verbal insults, presented to participants in an experiment like this? Our study took place in a lab setting where the statements were massively repeated and came from fictitious agents whose identity and fictitious opinions have no bearing on the life of the participant. This is a far cry from real life. And, of course, it had to be: research with genuinely malicious insults, in a social context in which they really make sense, and coming from people that participants really care about, is very difficult to defend. The artificial repetitiveness and grouping of speech act types, however, was a deliberate choice, because our interest was precisely in what insults would do under such repetitive conditions. Our results show that even under these highly unnatural conditions, and in the absence of a richly defined interpersonal arena, verbal insults still "get at you" and continue to do so over time, at least at some level.

But where exactly? Where in processing might the P2 and LPP insult effects observed in our study arise? The ALC model (Van Berkum, 2018, 2019; see **Figure 1**) suggests that a genuine verbal insult can be emotionally evocative at several levels in the language comprehension cascade: the social intention of the speaker (e.g., to express contempt), the referential intention (i.e., the description of you as an idiot), the stance displayed (e.g., aggressive), and the signs used (e.g., very negative and sometimes even taboo words). Because the speakers in our lab experiment are only fictitious, and because there is no real interpersonal arena (other than between experimenter and participant; Clark, 2006) it seems unlikely that our P2 and LPP effects emerge because the reader has computed a speaker's social intention or emotional stance. After all, although we hinted at a social context in our instructions, it is unlikely that our participants thought a *real* Daan or Paul was aggressively insulting people here and now. Unless an experimental situation is designed to have real interaction between real (or well-faked) interlocutors, effects at the level of social intention and speaker stance depend entirely on the imagination that participants are willing to engage in.

What about the referential intention or "situation model," the state of affairs described? Although the referential intention of an item like "Linda is horrible" is not fully defined in the lab, it may well still be an unpleasant thing to read, particularly if your name is Linda. Insulting descriptions featuring you can be experienced as slightly unpleasant even when the social intention of the speaker is *benign*, such as in teasing, playfighting, and the informal marking of intimacy amongst friends (Van Berkum, 2018)—if *that* is possible, why not in the lab? In our experiment, though, the ERP insult effects in the P2 and LPP latency range were insensitive to whether (somebody with) the participant('s name) or somebody else was being referred to in

a very negative way. Our ERP effects therefore probably do not reflect the emotional consequence of a fully computed referential intention. Because a few participants later reported that the name of the other person coincided with the name of somebody they knew, we cannot exclude that any potential referential effect was somewhat diluted by this. However, when we stick to the original question, whether negative description of the self vs. somebody else matters to ERPs and skin conductance, the answer is a clear "not reliably here."[6]

Turning to the level of the signs, the verbal insults in our experiment all use very negative words (e.g., "horrible," "ugly," "bitch," "disappointment," etcetera), signs that may simply deliver their stored emotional "payload" upon retrieval and attract more emotional attention as part of that, independent of the precise sentence-level message and the social intention of the speaker on this particular occasion (Van Berkum, 2018, 2019). We already offered this as an account for the fact that insults elicited a larger P2 than compliments and continued to do so all the way through the session. On a lexical account, this robustness makes sense, because the valence or other stable aspects of emotional meaning stored with words in long-term memory is usually the result of a lifetime of experience with how words are used in one's language. Just as one cannot completely overrule the conventional non-emotional aspects of the meaning of "liar," "ugly," "bitch" or "disappointment" in a single experimental session, it might be hard to completely undo the emotional part of meaning.

Where does the emotional meaning of these words come from? As laid out in more detail elsewhere (Van Berkum, 2019), it presumably reflects experiences that pair the word with emotions triggered by other things in a sufficiently reliable way, such as other emotion-laden words that it typically collocates with (e.g., Hauser and Schwarz, 2018; Snefjella et al., 2020), the typical social intention or emotional stance of the speaker using this word, the typical impact on listeners of this word, or the nature of the things referred to by this word. Because of these correlations, the release of that stored emotional meaning as part of lexical retrieval can be seen as a quick-and-dirty context-free prediction of what the word might in the end mean in this particular context. In the confines of an artificial lab experiment, those memory-based predictions will usually be wrong. After all, in the lab, the use of, say, the word "bitch" does not predict an angry or playfighting experimenter, and "disappointment" will usually not signal social rejection—they are just words flashed on the screen as part of some bits of language to attend to, for money or course credits. Nevertheless, these memory-based predictions will probably unavoidably be made. We suspect that the insult-induced P2 observed in our study reflects the automatic emotional attention that such memory-based predictions bring about.

Just like the P2 effect, the later insult-induced LPP effect is insensitive to whether the participant or someone else was involved. However, in contrast to the former, the latter does (in

---

[6]*Across speech act types,* statements referring to the self did elicit larger positivities in the LPP (and preceding N400-) latency range, as well as increased skin conductance responses (i.e., arousal), than statements referring to others; see the **Supplementary Material**. This confirms the effectiveness of the self-other manipulation.

the 350–450 ms latency range) significantly diminish as insults are repeated again and again. The most probable explanation is not that the effect has a different trigger, but that it reflects the decreased allocation of more costly resources to the same lexical trigger. Apart from how it contributes to higher levels of composite and inferential meaning, a very negative word by itself elicits a cascade of different processes, some of which are automatic, and others less so. With emotionally evocative words or pictures presented in isolation, this is in fact how the LPP is construed: an index of more controlled processing (Citron, 2012), or a "more prolonged and integrated conceptual analysis" (Frank and Sabatinelli, 2019, p. 9) that is part of a larger processing cascade initiated by the same functional trigger. Some have argued that this late process involves reappraisal (Hajcak and Nieuwenhuis, 2006; Hajcak et al., 2006; Moser et al., 2006; Herbert et al., 2011), a usually effortful attempt to regulate emotional responses by reinterpreting the original stimulus. Late ERP positivities have also been associated with conscious processing more generally (see Dehaene and King, 2016), i.e., as reflecting additional "global-workspace" processing that might involve reappraisal, but can also just index some other conscious, more effortful response to the input. It is easy to imagine our participants reappraising (e.g., "this is not real, I'm in a lab"), or to otherwise reflect on the offensive materials. Either way, an insult-induced LPP effect that reflects effortful conscious processing might as such also be more vulnerable to repetition, perhaps in part because of increasing fatigue and decreasing motivation. This would explain the reliable decline of this effect in the 350–450 ms latency range.

If all this is correct, why did the very negative words in our insults not elicit a larger skin conductance response than the positive words in compliments, at least in the first part of the session, where we also observed an insult-induced LPP effect? We simply do not know. In a repetition study using emotional and neutral pictures, Schupp et al. (2006) observed a comparable dissociation between emotion-induced ERP effects that were insensitive to repetition and skin conductance effects that collapsed with repetition, and explained this in terms of the functional relevance of continued cognitive processing but decreased action preparation. However, in the face of a robust skin conductance effect to neutral, factually correct statements that lasts all the way to the end of the session, we cannot easily invoke this account for our findings.

Does the fact that insults do not increase arousal relative to the other speech acts render the insult-induced ERP-effects "non-emotional"? Note that one of the core jobs of emotion is to have *cognitive* effects, such as increased attention, stronger memorization, and changes in motivational orientation (see Van Berkum, 2022, for review). It is therefore to be expected that emotion-relevant manipulations can modulate ERP components that show up in "cognitive" psycholinguistic research as well. Furthermore, what is emotional or not is not that easy to define. In the field of emotion research, for example, surprise is actually viewed as an emotion (and, in Basic Emotions theory, even as one of the six basic emotions, Ekman and Cordaro, 2011). Also, the LPP has recently been related to the P300 (Hajcak and Foti, 2020) not because the former should be considered as

irrelevant to emotion, but because both effects may index the same thing: significance to the participant. That significance may be artificially created because of task instructions, or may come naturally with the materials (Bradley, 2009). Either way, our most natural automatic response to significance—to the detection that stimuli somehow bear on our local or more general interests—is with the type of response that is the subject matter of emotion science (Van Berkum, 2022). Without claiming that the ERP components in our current study are uniquely emotional (let alone moral), we think it is safe to assume that their modulation by insulting materials has something to do with emotion.

In contrast to what we expected, though, insults in our study do not "keep their sting" across repetitions more than compliments keep their more benign emotional impact. P2 and SCR amplitudes declined at comparable rates for all speech acts, and insult-induced LPP amplitudes (spanning the 350–450 and 500–650 ms latency ranges) actually declined at a faster rate than compliment-induced ones. Also, there was no clear evidence that *who* was being insulted or complimented mattered to processing. We do not take our results to show that in real contextualized language use, people get used to insults as fast—or faster—than to compliments, nor that who is insulted does not matter. What is more likely is that, in lab experiments such as these, insults are "lexical mini-slaps" at best. Those are interesting to study, of course, and lexical effects may have a story to tell about memory traces and predictions involving emotions triggered at higher levels of real-life language use. However, in line with our ALC-model analysis, our data suggest that it takes more to study the impact of real verbal slaps in the face. In fact, perhaps more than current ethics standards allow.

## Limitations

The most important limitation of our study is exactly that: our insults will have been recognized as such, but as decontextualized statements, their insulting power is limited. Interestingly, in the exit interviews, some participants did actually report feeling insulted, perhaps also because of our occasional use of swearwords. But on the whole, the absence of a real communicative situation, i.e., a real speaker addressing you about things that matter, will have placed an upper bound on the impact of our materials. Although the use of audio- or video-recorded speakers might improve things to some extent, studying real insults in the lab will remain a challenge[7]. Unless insulting statements are embedded in real (or well-faked) social interaction, processing studies are more likely to pick up on memory traces reflecting earlier experiences with real insults that participants bring to the lab than on what a particular insult can really evoke in a person, *here and now*.

A few other limitations should also be kept in mind. First, participants did not have to do anything with the statements, other than to attend to them. If the ERP effects themselves signal involvement this is not necessarily problematic, but with

---

[7]Presumably, decontextualization is a bigger problem for processing research on the actual emotional, "perlocutionary" effects of insults (like ours) than for 'locution-oriented' semantic-pragmatic research with sentence acceptability ratings (such as Cepollaro et al., 2021).

null results—such as no impact of who is being insulted—a taskless design can become a vulnerability. Second, we presented our stimuli in repetitive blocks. This was done on purpose, to allow us to study the response of various speech acts to massive repetition, and to test the idea that the response to insults directed at the self would be more robust than other speech act types. However, blockwise presentation will undoubtedly also have led to expectations, such as that the next item will also feature your name. Third, we have not designed our study to unconfound the effects of repetition from that of the mere passage of time. The two are always confounded in real life, and our questions were not about general adaptation effects, or the effects of mere time on task. Fourth, we assumed that the use of the participant's name would bring the statements to bear on the participant's self. Although this is not an unreasonable assumption, and echoes assumptions made in research with personal pronouns ("I," "you," etc.), it is good to keep in mind that there is some indirectness involved here. Fifth, although we successfully manipulated self- vs. other-directedness in our statements, the occasional presence of other-person names coinciding with the name of somebody they knew may have reduced our power to detect a referentially conditioned effect somewhat. Of course, the issue *was* self vs. other, so in one way, the occasional presence of names of familiar others in fact strengthens the design. At the same time, though, future studies that specifically query a referentially dependent insult effect may be more successful if they avoid the inclusion of familiar other's names. And sixth, for practical reasons that concerned the construction of our materials, we only tested female participants. We have no particular reason to believe that the observed P2 and LPP effects are gender-specific. However, some of the insulting and complimenting predicates that we used are expected to work very differently with a different gender, something to be taken into account in future replication research.

## Conclusions

Genuine verbal insults violate the universal moral imperative not to harm other people. They are also threats to one's face or reputation, something that is not to be taken lightly in a species that depends on cooperation with ever changing partners. We showed that when compared to more positive evaluative statements, verbal insults elicited a very rapid P2 response that did not diminish as a function of massive repetition, and also did not depend on who was involved. The fact that verbal transgressions robustly capture attention within 250 ms after they occur, in an undefined laboratory situation where no real interaction occurs, is not only indicative of our sensitivity to undesirable social behavior, but also in line with the idea that the evaluation of such behavior is to some extent automatic. For members of an ultra-social species that depend on ever-changing cooperation, reputation, respect, and interpersonal trust, getting slapped in the face, or seeing somebody else suffer that fate, is a highly salient event, regardless of the precise context. It should be no surprise that the verbal equivalent is similarly evocative, and that the reflexive, memory-based traces of this can even be found in a psycholinguistic experiment that does not involve natural interpersonal interaction.

## DATA AVAILABILITY STATEMENT

All information regarding the experimental set-up, data, and analysis of the experiment can be found in our online repository: https://doi.org/10.24416/UU01-I0U3TW.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Linguistics Chamber of the Faculty of Humanities Ethics assessment Committee at Utrecht University. The participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

Authors contributed equally to all aspects of the research, except for testing (only MS and HDM) and analysis (only MS). All authors contributed to the article and approved the submitted version.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2022.910023/full#supplementary-material

## REFERENCES

Baumeister, R. F., and Bushman, B. J. (2021). *Social Psychology and Human Nature*. Boston, MA: Cengage.

Bayer, M., Sommer, W., and Schacht, A. (2010). Reading emotional words within sentences: the impact of arousal and valence on event-related potentials. *Int. J. Psychophysiol.* 78, 299–307. doi: 10.1016/j.ijpsycho.2010.09.004

Benau, E. M., and Atchley, R. A. (2020). Some compliments (and insults) are more heartfelt. *J. Psychophysiol.* 34, 50–59. doi: 10.1027/0269-8803/a000235

Benedek, M., and Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *J. Neurosci. Methods* 190, 80–91. doi: 10.1016/j.jneumeth.2010.04.028

Bernat, E., Bunce, S., and Shevrin, H. (2001). Event-related brain potentials differentiate positive and negative mood adjectives during both supraliminal and subliminal visual processing. *Int. J. Psychophysiol.* 42, 11–34. doi: 10.1016/S0167-8760(01)00133-7

Boucsein, W. (2012). *Electrodermal Activity*. New York, NY: Springer. doi: 10.1007/978-1-4614-1126-0

Bradley, M. M. (2009). Natural selective attention: orienting and emotion. *Psychophysiology* 46, 1–11. doi: 10.1111/j.1469-8986.2008.00702.x

Brouwer, H., Crocker, M. W., Venhuizen, N. J., and Hoeks, J. C. (2017). A neurocomputational model of the N400 and the P600 in language processing. *Cogn. Sci.* 41, 1318–1352. doi: 10.1111/cogs.12461

Brown, P., and Levinson, S. C. (1987). *Politeness: Some Universals in Language Usage.* Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511813085

Brummelman, E., Nelemans, S. A., Thomaes, S., and Orobio de Castro, B. (2017). When parents' praise inflates, children's self-esteem deflates. *Child Dev.* 88, 1799–1809. doi: 10.1111/cdev.12936

Cacioppo, J. T., Crites, S. L., Berntson, G. G., and Coles, M. G. H. (1993). If attitudes affect how stimuli are processed, should they not affect the event-related brain potential? *Psychol. Sci.* 4, 108–112. doi: 10.1111/j.1467-9280.1993.tb00470.x

Cacioppo, J. T., Crites, S. L., Gardner, W. L., and Berntson, G. G. (1994). Bioelectrical echoes from evaluative categorizations: I. A late positive brain potential that varies as a function of trait negativity and extremity. *J. Pers. Soc. Psychol.* 67, 115–125. doi: 10.1037/0022-3514.67.1.115

Cacioppo, J. T., Larsen, J. T., Smith, N. K., and Berntson, G. G. (2004). "The affect system: what lurks below the surface of feelings?" in *Feelings and Emotions: The Amsterdam Conference*, eds A. S. R. Manstead, N. H. Frijda, and A. H. Fischer (New York, NY: Cambridge University Press), 223–242. doi: 10.1017/CBO9780511806582.014

Campbell-Sills, L., and Stein, M. B. (2007). Psychometric analysis and refinement of the Connor-Davidson Resilience Scale (CD-RISC): validation of a 10-item measure of resilience. *J. Trauma. Stress* 20, 1019–1028. doi: 10.1002/jts.20271

Carretié, L. (2014). Exogenous (automatic) attention to emotional stimuli: a review. *Cogn. Affect. Behav. Neurosci.* 14, 1228–1258. doi: 10.3758/s13415-014-0270-2

Carretié, L., Hinojosa, J. A., Albert, J., López-Martín, S., De La Gándara, B. S., Igoa, J. M., et al. (2008). Modulation of ongoing cognitive processes by emotionally intense words. *Psychophysiology* 45, 188–196. doi: 10.1111/j.1469-8986.2007.00617.x

Carretié, L., Mercado, F., Tapia, M., and Hinojosa, J. A. (2001). Emotion, attention, and the 'negativity bias', studied through event-related potentials. *Int. J. Psychophysiol.* 41, 75–85. doi: 10.1016/S0167-8760(00)00195-1

Cepollaro, B., Domaneschi, F., and Stojanovic, I. (2021). When is it ok to call someone a jerk? An experimental investigation of expressives. *Synthese* 198, 9273–9292. doi: 10.1007/s11229-020-02633-z

Citron, F. M. (2012). Neural correlates of written emotion word processing: a review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain Lang.* 122, 211–226. doi: 10.1016/j.bandl.2011.12.007

Citron, F. M., Weekes, B. S., and Ferstl, E. C. (2013). Effects of valence and arousal on written word recognition: time course and ERP correlates. *Neurosci. Lett.* 533, 90–95. doi: 10.1016/j.neulet.2012.10.054

Clark, H. H. (2006). "Social actions, social commitments," in *Roots of Human Sociality: Culture, Cognition and Interaction*, eds N. J. Enfield and S. C. Levinson (Oxford: Berg), 126–150. doi: 10.4324/9781003135517-6

Codispoti, M., Bradley, M. M., and Lang, P. J. (2001). Affective reactions to briefly presented pictures. *Psychophysiology* 38, 474–478. doi: 10.1111/1469-8986.3830474

Damasio, A. (2010). *Self Comes to Mind: Constructing the Conscious Brain.* New York, NY: Pantheon/Random House.

Davidson, R. J., and Begley, S. (2013). *The Emotional Life of your Brain.* New York, NY: Hudson Street Press.

Davis, M. H. (1983). Measuring individual differences in empathy: evidence for a multidimensional approach. *J. Pers. Soc. Psychol.* 44, 113–126. doi: 10.1037/0022-3514.44.1.113

Dawson, M. E., Schell, A. M., and Filion, D. L. (2007). "The electrodermal system," in *Handbook of Psychophysiology*, eds J. T. Cacioppo, L. G. Tassinary, and G. Berntson (Cambridge University Press), 159–181.

Dehaene, S., and King, J. (2016). "Decoding the dynamics of conscious perception: the temporal generalization method," in *Micro-, Meso-and Macro-Dynamics of the Brain*, eds G. Buzsáki and Y. Christen (Cham: Springer), 85–97. doi: 10.1007/978-3-319-28802-4_7

Delplanque, S., Lavoie, M. E., Hot, P., Silvert, L., and Sequeira, H. (2004). Modulation of cognitive processing by emotional valence studied through event-related potentials in humans. *Neurosci. Lett.* 356, 1–4. doi: 10.1016/j.neulet.2003.10.014

Eilola, T. M., and Havelka, J. (2011). Behavioural and physiological responses to the emotional and taboo Stroop tasks in native and non-native speakers of English. *Int. J. Biling.* 15, 353–369. doi: 10.1177/1367006910379263

Ekman, P., and Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion Review* 3, 364–370. doi: 10.1177/1754073911410740

Enfield, N. J. (2013). *Relationship Thinking: Agency, Enchrony, and Human Sociality.* New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780199338733.001.0001

Espuny, J., Jiménez-Ortega, L., Casado, P., Fondevila, S., Muñoz, F., Hernández-Gutiérrez, D., et al. (2018). Event-related brain potential correlates of words' emotional valence irrespective of arousal and type of task. *Neurosci. Lett.* 670, 83–88. doi: 10.1016/j.neulet.2018.01.050

Fields, E. C., and Kuperberg, G. R. (2012). It's all about you: An ERP study of emotion and self-relevance in discourse. *NeuroImage.* 62, 562–574. doi: 10.1016/j.neuroimage.2012.05.003

Fields, E. C., and Kuperberg, G. R. (2015). Loving yourself more than your neighbor: ERPs reveal online effects of a self-positivity bias. *Soc. Cogn. Affect. Neurosci.* 10, 1202–1209. doi: 10.1093/scan/nsv004

Fields, E. C., and Kuperberg, G. R. (2016). Dynamic effects of self-relevance and task on the neural processing of emotional words in context. *Front. Psychol.* 6:2003. doi: 10.3389/fpsyg.2015.0200

Foti, D., Hajcak, G., and Dien, J. (2009). Differentiating neural responses to emotional pictures: evidence from temporal-spatial PCA. *Psychophysiology* 46, 521–530. doi: 10.1111/j.1469-8986.2009.00796.x

Foucart, A., Moreno, E., Martin, C. D., and Costa, A. (2015). Integration of moral values during L2 sentence processing. *Acta Psychol.* 162, 1–12. doi: 10.1016/j.actpsy.2015.09.009

Frank, D. W., and Sabatinelli, D. (2019). Hemodynamic and electrocortical reactivity to specific scene contents in emotional perception. *Psychophysiology.* 56:e13340.doi: 10.1111/psyp.13340

Frijda, N. H. (2007). *The laws of emotion.* Lawrence Erlbaum Associates Publishers.

Gantman, A. P., and Van Bavel, J. J. (2015). Moral perception. *Trends Cogn. Sci.* 19, 631–633. doi: 10.1016/j.tics.2015.08.004

Goffman, E. (1967). *Interaction ritual: Essays on Face-to-Face Interaction.* Oxford: Aldine.

Greene, J. (2013). *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them.* New York, NY: Penguin Press.

Haidt, J. (2012). *The Righteous Mind: Why Good People Are Divided by Politics and Religion.* New York, NY: Pantheon/Random House.

Hajcak, G., and Foti, D. (2020). Significance? significance! empirical, methodological, and theoretical connections between the late positive potential and P300 as neural responses to stimulus significance: an integrative review. *Psychophysiology* 57, e13570. doi: 10.1111/psyp.13570

Hajcak, G., MacNamara, A., Foti, D., Ferri, J., and Keil, A. (2013). The dynamic allocation of attention to emotion: simultaneous and independent evidence from the late positive potential and steady state visual evoked potentials. *Biol. Psychol.* 92, 447–455. doi: 10.1016/j.biopsycho.2011.11.012

Hajcak, G., Moser, J. S., and Simons, R. F. (2006). Attending to affect: appraisal strategies modulate the electrocortical response to arousing pictures. *Emotion* 6, 517–522. doi: 10.1037/1528-3542.6.3.517

Hajcak, G., and Nieuwenhuis, S. (2006). Reappraisal modulates the electrocortical response to unpleasant pictures. *Cogn. Affect. Behav. Neurosci.* 6, 291–297. doi: 10.3758/CABN.6.4.291

Hajcak, G., Weinberg, A., MacNamara, A., and Foti, D. (2012). "ERPs and the study of emotion," in *The Oxford Handbook of Event-Related Potential Components*, eds S. J. Luck and E. S. Kappenman (Oxford: Oxford University Press), 441–472. doi: 10.1093/oxfordhb/9780195374148.013.0222

Halgren, E., Dhond, R., Christensen, N., van Petten, C., Marinkovic, K., Lewine, J., et al. (2002). N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *Neuroimage* 17, 1101–1116. doi: 10.1006/nimg.2002.1268

Harris, C. L. (2004). Bilingual speakers in the lab: psychophysiological measures of emotional reactivity. *J. Multiling. Multicult. Dev.* 25, 223–247. doi: 10.1080/01434630408666530

Harris, C. L., Ayçíçegí, A., and Gleason, J. B. (2003). Taboo words and reprimands elicit greater autonomic reactivity in a first language than in a second language. *Appl. Psycholinguist.* 24, 561–579. doi: 10.1017/S0142716403000286

Hauser, D. J., and Schwarz, N. (2018). How seemingly innocuous words can bias judgment: semantic prosody and impression formation. *J. Exp. Soc. Psychol.* 75, 11–18. doi: 10.1016/j.jesp.2017.10.012

Herbert, C., Junghofer, M., and Kissler, J. (2008). Event related potentials to emotional adjectives during reading. *Psychophysiology* 45, 487–498. doi: 10.1111/j.1469-8986.2007.00638.x

Herbert, C., Kissler, J., Junghöfer, M., Peyk, P., and Rockstroh, B. (2006). Processing of emotional adjectives: evidence from startle EMG and ERPs. *Psychophysiology* 43, 197–206. doi: 10.1111/j.1469-8986.2006.00385.x

Herbert, C., Pauli, P., and Herbert, B. M. (2011). Self-reference modulates the processing of emotional stimuli in the absence of explicit self-referential appraisal instructions. *Soc. Cogn. Affect. Neurosci.* 6, 653–661. doi: 10.1093/scan/nsq082

Holt, D. J., Lynn, S. K., and Kuperberg, G. R. (2009). Neurophysiological correlates of comprehending emotional meaning in context. *J. Cogn. Neurosci.* 21, 2245–2262. doi: 10.1162/jocn.2008.21151

Huang, Y., and Luo, Y. (2006). Temporal course of emotional negativity bias: an ERP study. *Neurosci. Lett.* 398, 91–96. doi: 10.1016/j.neulet.2005.12.074

Hundrieser, M., and Stahl, J. (2016). How attitude strength and information influence moral decision making: evidence from event-related potentials. *Psychophysiology* 53, 678–688. doi: 10.1111/psyp.12599

Irvine, W. B. (2013). *A Slap in the Face: How to Live in a World of Snubs, Put-downs, and Insults*. New York, NY: Oxford University Press.

Ito, T. A., Larsen, J. T., Smith, N. K., and Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: the negativity bias in evaluative categorizations. *J. Pers. Soc. Psychol.* 75, 887–900. doi: 10.1037/0022-3514.75.4.887

Jackendoff, R. (2007). A parallel architecture perspective on language processing. *Brain Res.* 1146, 2–22. doi: 10.1016/j.brainres.2006.08.111

Jay, T. (2009). The utility and ubiquity of taboo words. *Perspect. Psychol. Sci.* 4, 153–161. doi: 10.1111/j.1745-6924.2009.01115.x

Kanske, P., and Kotz, S. A. (2007). Concreteness in emotional words: ERP evidence from a hemifield study. *Brain Res*. 1148, 138–148. doi: 10.1016/j.brainres.2007.02.044

Keuleers, E., Brysbaert, M., and New, B. (2010). SUBTLEX-NL: a new measure for Dutch word frequency based on film subtitles. *Behav. Res. Methods* 42, 643–650. doi: 10.3758/BRM.42.3.643

Kisley, M. A., Wood, S., and Burrows, C. L. (2007). Looking at the sunny side of life: age-related change in an event-related potential measure of the negativity bias. *Psychol. Sci.* 18, 838–843. doi: 10.1111/j.1467-9280.2007.01988.x

Kissler, J., Assadollahi, R., and Herbert, C. (2006). Emotional and semantic networks in visual word processing: insights from ERP studies. *Prog. Brain Res.* 156, 147–183. doi: 10.1016/S0079-6123(06)56008-X

Kissler, J., Herbert, C., Winkler, I., and Junghofer, M. (2009). Emotion and attention in visual word processing—an ERP study. *Biol. Psychol.* 80, 75–83. doi: 10.1016/j.biopsycho.2008.03.004

Kunkel, A., Filik, R., Mackenzie, I. G., and Leuthold, H. (2018). Task-dependent evaluative processing of moral and emotional content during comprehension: an ERP study. *Cogn. Affect. Behav. Neurosc.* 18, 389–409. doi: 10.3758/s13415-018-0577-5

Kutas, M., and Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn. Sci.* 4, 463–470. doi: 10.1016/S1364-6613(00)01560-6

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123

Lang, P. J., Greenwald, M. K., Bradley, M. M., and Hamm, A. O. (1993). Looking at pictures: affective, facial, visceral, and behavioral reactions. *Psychophysiology* 30, 261–273. doi: 10.1111/j.1469-8986.1993.tb03352.x

Langeslag, S. J., and Van Strien, J. W. (2018). Early visual processing of snakes and Langeslag and Van Strien, 2018;angry faces: an ERP study. *Brain Res.* 1678, 297–303. doi: 10.1016/j.brainres.2017.10.031

Lau, E. F., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nat. Rev. Neurosci.* 9, 920–933. doi: 10.1038/nrn2532

Lazarus, J. (2021). Negativity bias: An evolutionary hypothesis and an empirical programme. *Learn. Motivat.* 75:101731. doi: 10.1016/j.lmot.2021.101731

Leary, M. R. (2005). Sociometer theory and the pursuit of relational value: getting to the root of self-esteem. *Eur. Rev. Soc. Psychol.* 16, 75–111. doi: 10.1080/10463280540000007

Lei, Y., Dou, H., Liu, Q., Zhang, W., Zhang, Z., and Li, H. (2017). Automatic processing of emotional words in the absence of awareness: the critical role of P2. *Front. Psychol.* 8, 592. doi: 10.3389/fpsyg.2017.00592

Leuthold, H., Kunkel, A., Mackenzie, I. G., and Filik, R. (2015). Online processing of moral transgressions: ERP evidence for spontaneous evaluation. *Soc. Cogn. Affect. Neurosci.* 10, 1021–1029. doi: 10.1093/scan/nsu151

Luck, S. J., and Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology* 54, 146–157. doi: 10.1111/psyp.12639

Luck, S. J., and Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology* 31, 291–308. doi: 10.1111/j.1469-8986.1994.tb02218.x

MacDonald, G., and Leary, M. R. (2005). Why does social exclusion hurt? The relationship between social and physical pain. *Psychol. Bull.* 131, 202–223. doi: 10.1037/0033-2909.131.2.202

Moser, J. S., Hajcak, G., Bukay, E., and Simons, R. F. (2006). Intentional modulation of emotional responding to unpleasant pictures: an ERP study. *Psychophysiology* 43, 292–296. doi: 10.1111/j.1469-8986.2006.00402.x

Otten, M., Mann, L., van Berkum, J. J. A., and Jonas, K. J. (2017). No laughing matter: how the presence of laughing witnesses changes the perception of insults. *Soc. Neurosci.* 12, 182–193. doi: 10.1080/17470919.2016.1162194

Otten, M., and Van Berkum, J. J. A. (2007). What makes a discourse constraining? Comparing the effects of discourse message and scenario fit on the discourse-dependent N400 effect. *Brain Res.* 1153, 166–177. doi: 10.1016/j.brainres.2007.03.058

Panksepp, J., and Biven, L. (2012). *The Archaeology of Mind: Neuroevolutionary Origins of Human Emotions*. New York, NY: W. W. Norton.

Peng, X., Jiao, C., Cui, F., Chen, Q., Li, P., and Li, H. (2017). The time course of indirect moral judgment in gossip processing modulated by different agents. *Psychophysiology* 54, 1459–1471. doi: 10.1111/psyp.12893

Pessoa, L. (2013). "The impact of emotion on cognition," in *The Oxford Handbook of Cognitive Neuroscience: Volume 2: The Cutting Edges*, eds K. N. Ochsner and S. Kosslyn, (New York, NY: Oxford University Press).

Polich, J. (2012). "Neuropsychology of P300," in *The Oxford Handbook of Event-Related Potential Components*, eds S. J. Luck, and E. S. Kappenman (Oxford: Oxford University Press), 159–188.

Quené, H., and van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: a tutorial. *Speech Commun.* 43, 103–121. doi: 10.1016/j.specom.2004.02.004

Quené, H., and van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *J. Mem. Lang.* 59, 413–425. doi: 10.1016/j.jml.2008.02.002

Recio, G., Conrad, M., Hansen, L. B., and Jacobs, A. M. (2014). On pleasure and thrill: the interplay between arousal and valence during visual word recognition. *Brain Lang.* 134, 34–43. doi: 10.1016/j.bandl.2014.03.009

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

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

Rosenberg, M. (1965). *Society and the Adolescent Self-Image*. Princeton, NJ: Princeton University Press.

Rozin, P., and Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Pers. Soc. Psychol. Rev.* 5, 296–320. doi: 10.1207/S15327957PSPR0504_2

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

Sabatinelli, D., Lang, P. J., Keil, A., and Bradley, M. M. (2007). Emotional perception: correlation of functional MRI and event-related potentials. *Cereb. Cortex* 17, 1085–1091. doi: 10.1093/cercor/bhl017

Schupp, H. T., Cuthbert, B. N., Bradley, M. M., Cacioppo, J. T., Ito, T., and Lang, P. J. (2000). Affective picture processing: the late positive potential is modulated by motivational relevance. *Psychophysiology*. 37, 257–261.

Schupp, H. T., Flaisch, T., Stockburger, J., and Junghöfer, M. (2006). Emotion and attention: event-related brain potential studies. *Prog. Brain Res*. 156, 31–51. doi: 10.1016/S0079-6123(06)56002-9

Schupp, H. T., Junghöfer, M., Weike, A. I., and Hamm, A. O. (2004). The selective processing of briefly presented affective pictures: an ERP analysis. *Psychophysiology* 41, 441–449. doi: 10.1111/j.1469-8986.2004.00174.x

Schupp, H. T., and Kirmse, U. M. (2021). Case-by-case: emotional stimulus significance and the modulation of the EPN and LPP. *Psychophysiology* 58, e13766. doi: 10.1111/psyp.13766

Scott, G. G., O'Donnell, P. J., Leuthold, H., and Sereno, S. C. (2009). Early emotion word processing: evidence from event-related potentials. *Biol. Psychol*. 80, 95–104. doi: 10.1016/j.biopsycho.2008.03.010

Seasholtz, M. B., and Kowalski, B. (1993). The parsimony principle applied to multivariate calibration. *Anal. Chim. Acta* 277, 165–177. doi: 10.1016/0003-2670(93)80430-S

Shin, Y. S., and Niv, Y. (2021). Biased evaluations emerge from inferring hidden causes. *Nat. Hum. Behav*. 5, 1180–1189. doi: 10.1038/s41562-021-01065-0

Siakaluk, P. D., Pexman, P. M., Dalrymple, H. R., Stearns, J., and Owen, W. J. (2011). Some insults are more difficult to ignore: the embodied insult Stroop effect. *Lang. Cogn. Process*. 26, 1266–1294. doi: 10.1080/01690965.2010.521021

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., and Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature* 439, 466–469. doi: 10.1038/nature04271

Smith, N. K., Cacioppo, J. T., Larsen, J. T., and Chartrand, T. L. (2003). May I have your attention, please: electrocortical responses to positive and negative stimuli. *Neuropsychologia* 41, 171–183. doi: 10.1016/S0028-3932(02)00147-1

Snefjella, B., Lana, N., and Kuperman, V. (2020). How emotion is learned: semantic learning of novel words in emotional contexts. *J. Mem. Lang*. 115, 104171. doi: 10.1016/j.jml.2020.104171

Spielberger, C., Gorsuch, R. L., Lushene, R., Vagg, P., and Jacobs, G. (1983). *Manual for the State-Trait Anxiety Scale*. Palo Alto, CA: Consulting Psychologists Press.

Swaab, T. Y., Ledoux, K., Camblin, C. C., and Boudewyn, M. A. (2012). "Language related ERP components," in *Oxford Handbook of Event-Related Potential Components*, eds S. J. Luck and E. S. Kappenman (New York, NY: Oxford University Press), 397–440. doi: 10.1093/oxfordhb/9780195374148.013.0197

Tennie, C., Frith, U., and Frith, C. D. (2010). Reputation management in the age of the world-wide web. *Trends Cogn. Sci*. 14, 482–488. doi: 10.1016/j.tics.2010.07.003

Tomasello, M. (2008). *Origins of Human Communication*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/7551.001.0001

Tomasello, M., and Herrmann, E. (2010). Ape and human cognition: what's the difference? *Curr. Direct. Psychol. Res*. 19, 3–8. doi: 10.1177/0963721409359300

Trauer, S. M., Andersen, S. K., Kotz, S. A., and Müller, M. M. (2012). Capture of lexical but not visual resources by task-irrelevant emotional words: a combined ERP and steady-state visual evoked potential study. *Neuroimage* 60, 130–138. doi: 10.1016/j.neuroimage.2011.12.016

Unkelbach, C., Koch, A., and Alves, H. (2021). Explaining negativity dominance without processing bias. *Trends Cogn. Sci*. 25, 429–430. doi: 10.1016/j.tics.2021.04.005

Van Berkum, J. J. A. (2009). "The neuropragmatics of 'simple' utterance comprehension: An ERP review" in *Semantics and Pragmatics: From Experiment to Theory* (Basingstoke: Palgrave Macmillan), 276–316.

Van Berkum, J. J. A. (2012). "The electrophysiology of discourse and conversation" in *The Cambridge Handbook of Psycholinguistics*, eds M. J. Spivey, K.

McRae, and M. F. Joanisse (New York, NY: Cambridge University Press), 589–614.

Van Berkum, J. J. A. (2018). "Language comprehension, emotion, and sociality: aren't we missing something?" in *The Oxford Handbook of Psycholinguistics*, eds S. Rueschemeyer and G. Gaskell (Oxford: Oxford University Press), 644–669. doi: 10.1093/oxfordhb/9780198786825.013.28

Van Berkum, J. J. A. (2019). "Language comprehension and emotion: where are the interfaces, and who cares?" in *Oxford Handbook of Neurolinguistics*, eds G. de Zubicaray and N. O. Schiller (Oxford: Oxford University Press), 736–766. doi: 10.1093/oxfordhb/9780190672027.013.29

Van Berkum, J. J. A. (2022). "A survey of emotion theories and their relevance to language research," in *Language and Emotion: An International Handbook*, eds G. L. Schiewer, J. Altarriba, and B. C. Ng (Berlin: Mouton de Gruyter).

Van Berkum, J. J. A., Holleman, B. C., Nieuwland, M. S., Otten, M., and Murre, J. M. J. (2009). Right or wrong? The brain's fast response to morally objectionable statements. *Psychol. Sci*. 20, 1092–1099. doi: 10.1111/j.1467-9280.2009.02411.x

Vieitez, L., Haro, J., Ferré, P., Padrón, I., and Fraga, I. (2021). Unraveling the mystery about the negative valence bias: does arousal account for processing differences in unpleasant words? *Front. Psychol*. 12, 748726. doi: 10.3389/fpsyg.2021.748726

Vuilleumier, P. (2005). How brains beware: neural mechanisms of emotional attention. *Trends Cogn. Sci*. 9, 585–594. doi: 10.1016/j.tics.2005.10.011

Wabnitz, P., Martens, U., and Neuner, F. (2012). Cortical reactions to verbal abuse: event-related brain potentials reflecting the processing of socially threatening words. *Neuroreport* 23, 774–779. doi: 10.1097/WNR.0b013e328356f7a6

Wabnitz, P., Martens, U., and Neuner, F. (2016). Written threat: electrophysiological evidence for an attention bias to affective words in social anxiety disorder. *Cogn. Emot*. 30, 516–538. doi: 10.1080/02699931.2015.1019837

Wang, L., and Bastiaansen, M. (2014). Oscillatory brain dynamics associated with the automatic processing of emotion in words. *Brain Lang*. 137, 120–129. doi: 10.1016/j.bandl.2014.07.011

Weimer, N. R., Clark, S. L., and Freitas, A. L. (2019). Distinct neural responses to social and semantic violations: an N400 study. *Int. J. Psychophysiol*. 137, 72–81 B. (2020). *Statistics for Linguists: An Introduction Using R*. New York, NY; London: Routledge. doi: 10.4324/9781315165547jpsycho.2018.12.006

Winter, B. (2019). *Statistics for Linguists: An Introduction Using R*. London: Routledge.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.