



Task-dependent evaluative processing of moral and emotional content during comprehension: An ERP study

Angelika Kunkel¹ · Ruth Filik² · Ian Grant Mackenzie¹ · Hartmut Leuthold¹

Published online: 6 March 2018
© Psychonomic Society, Inc. 2018

Abstract

Recently, we showed that when participants passively read about moral transgressions (e.g., adultery), they implicitly engage in the evaluative (good–bad) categorization of incoming information, as indicated by a larger event-related brain potential (ERP) positivity to immoral than to moral scenarios (Leuthold, Kunkel, Mackenzie, & Filik in *Social, Cognitive, and Affective Neuroscience*, 10, 1021–1029, 2015). Behavioral and neuroimaging studies indicated that explicit moral tasks prioritize the semantic-cognitive analysis of incoming information but that implicit tasks, as used in Leuthold et al. (*Social, Cognitive, and Affective Neuroscience*, 10, 1021–1029, 2015), favor their affective processing. Therefore, it is unclear whether an affective categorization process is also involved when participants perform explicit moral judgments. Thus, in two experiments, we used similarly constructed morality and emotion materials for which their moral and emotional content had to be inferred from the context. Target sentences from negative vs. neutral emotional scenarios and from moral vs. immoral scenarios were presented using rapid serial visual presentation. In Experiment 1, participants made moral judgments for moral materials and emotional judgments for emotion materials. Negative compared to neutral emotional scenarios elicited a larger posterior ERP positivity (LPP) about 200 ms after critical word onset, whereas immoral compared to moral scenarios elicited a larger anterior negativity (500–700 ms). In Experiment 2, where the same emotional judgment to both types of materials was required, a larger LPP was triggered for both types of materials. These results accord with the view that morality scenarios trigger a semantic-cognitive analysis when participants explicitly judge the moral content of incoming linguistic information but an affective evaluation when judging their emotional content.

Keywords Moral judgment · Emotion judgment · Affective evaluation · LPP · Anterior negativity

We often find ourselves in situations in which a person is violating a prevailing social norm or moral value. For instance, if we find out that someone cheats in an exam or is telling a lie, we tend to spontaneously judge such behavior as bad or immoral. This reflects a fundamental aspect of human moral cognition, and it has been proposed that such judgments are based on affective or intuitive processes (Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Haidt, 2001). In line with this view, a recent event-related brain potential (ERP) study demonstrated that when participants passively read about everyday moral transgressions, they implicitly categorize the

described behaviors as good or bad as early as about 320 ms after the presentation of the critical word (cf. Leuthold, Kunkel, Mackenzie, & Filik, 2015), which was argued to be reflected by a late posterior positivity (LPP). However, there is also evidence in the literature suggesting that the semantic-cognitive analysis of incoming information dominates when explicit morality judgments are required, whereas the affective analysis is prioritized when emotion judgments are demanded (cf. Lai, Hagoort, & Casasanto, 2012; Sevinc & Spreng, 2014). Therefore, it remains unclear whether such a rapid evaluation process, as indicated by the LPP during a passive reading task, would, indeed, be found for explicit moral judgments as well. In addition, we aim to provide further support for the proposal that the LPP reflects affective processing of incoming information by also investigating whether an LPP is elicited by similarly constructed everyday emotional scenarios without a moral component. To this end, we will record ERPs that are elicited by scenarios describing moral transgressions and emotional events, to see whether, and to what extent, cognitive and affective processes

✉ Angelika Kunkel
angelika.kunkel@uni-tuebingen.de

¹ Department of Psychology, Eberhard Karls University of Tübingen, Schleichstrasse 4, 72076 Tübingen, Germany

² School of Psychology, University of Nottingham, University Park, Nottingham NG7 2RD, UK

are involved during discourse comprehension when explicit moral or emotional judgments are required.

Van Berkum, Holleman, Nieuwland, Otten, and Murre (2009) are, to our knowledge, the first to use a text comprehension approach to reveal the ERP correlates of moral cognition. Specifically, they investigated whether and how rapidly an individual's values influence the online linguistic meaning analysis of moral statements when explicit judgments were required. Male participants with two opposing value systems (members of a Dutch strict-Christian party vs. voters of parties with opposite moral-ethical programs, referred to here as non-Christians) were asked to rate their agreement with critical statements such as, "If my child were homosexual, I'd find this hard/easy to accept." The individual words forming these statements were presented using rapid serial visual presentation (RSVP), affording the measurement of immediate ERP responses to the critical word. They found that value-inconsistent compared to value-consistent critical words (e.g., *easy* vs. *hard* for strict-Christians and *hard* vs. *easy* for non-Christians, respectively) initially elicited a larger, broadly distributed positivity between 200 ms and 250 ms (P200), followed by a larger centroparietal negativity between 375 ms and 425 ms (N400), and finally a larger LPP between 500 ms and 650 ms.

N400 amplitude has been shown to respond to the predictability of a word within a given context (e.g., Kutas & Hillyard, 1984), to semantic anomalies at the discourse level (Van Berkum, Hagoort, & Brown, 1999), as well as to violations of world knowledge (e.g., Filik & Leuthold, 2008; Hagoort, Hald, Bastiaansen, & Petersson, 2004), reflecting the demands of meaning construction (for a review, see Kutas & Federmeier, 2011). Thus, Van Berkum et al. (2009) interpreted their N400 findings as indicating that readers immediately and automatically evaluate incoming information with respect to their personally held values, giving rise to a rapid value-based influence on meaning construction. They further speculated that this N400 effect overlaps with that of a single sustained ERP positivity that has an earlier onset than the (overlapping) N400, therefore emerging as a larger P200 and LPP for value-inconsistent than for value-consistent statements. Van Berkum and colleagues ruled out a cognitive, decision-related account of this LPP effect for two reasons. Firstly, it has been demonstrated that self-referential (true vs. false) statements that are unrelated to a person's value system do not elicit such an effect (Fischler, Bloom, Childers, Arroyo, & Perry, 1984). Secondly, negative compared to positive and neutral stimuli tend to elicit a larger LPP—reflecting a negativity bias in affective processing (e.g., Ito, Larsen, Smith, & Cacioppo, 1998)—even in language studies where participants either merely read for comprehension or made explicit decisions to critical emotion words (e.g., Holt, Lynn, & Kuperberg, 2009). Therefore, Van Berkum et al. took their LPP effect to reflect the automatic activation of the affect system, in accord with the view that the LPP relates to the

implicit evaluative processing of motivationally salient stimuli (cf. Hajcak, MacNamara, & Olvet, 2010).

However, in Van Berkum et al.'s (2009) study, the values held by the participants may have constrained their (implicit) expectations regarding the likely sentence endings. For instance, when persons holding strict Christian values read a statement (taken from their Table 1) beginning with "In a bad marriage, divorce is an . . .", based on their personal beliefs, they would not expect it to be continued with the word *acceptable*. Hence, similar to N400 effects driven by discourse-based or world-knowledge-based expectations (e.g., Filik & Leuthold, 2008; Hagoort et al., 2004; Van Berkum et al., 1999), it is conceivable that the larger N400 elicited by value-inconsistent than by value-consistent statements reflects an (implicit) emotional congruity effect that depends on the relation between the emotional features of the preceding context and the critical word. Crucially, a larger N400 to emotion words that were incongruent rather than congruent with the preceding context has been shown not only in studies using sequential prime-target tasks (e.g., Eder, Leuthold, Rothermund, & Schweinberger, 2012; Morris, Squires, Taber, & Lodge, 2003; Zhang, Lawson, Guo, & Jiang, 2006; but see Herring, Taylor, White, & Crites, 2011) but also in discourse comprehension studies using strongly constraining emotional contexts—for instance, when someone is described as being happy in a context that outlines either a positive or a negative event (e.g., León, Díaz, de Vega, & Hernández, 2010; Leuthold, Filik, Mackenzie, & Murphy, 2012). Accordingly, the N400 effect might reflect the more intense lexical or semantic processing for incongruent than for congruent moral statements, that is, a morality-unspecific language-related effect. If this conjecture would hold true, then Van Berkum and colleagues' interpretation of the P200 and LPP effect in terms of an affective evaluation of statements could be challenged as well. That is, the P200 effect reported might be attributed to the enhanced visual processing of incongruent or very unexpected linguistic inputs (e.g., Bohan, Leuthold, Hijikata, & Sanford, 2012; Ferretti, Singer, & Patterson, 2008; Leuthold et al., 2015), and the larger LPP following incongruent statements might reflect a P600-like semantic effect that is found in response to various types of semantic anomalies (for a review, see Kuperberg, 2007) and has been related to a continued re-analysis of linguistic input following a semantic processing conflict (cf. Kuperberg, 2007; Van de Meerendonk, Kolk, Chwilla, & Vissers, 2009).

A recent text comprehension study by Leuthold et al. (2015) used a different approach to examine the implicit rather than explicit evaluative processing of everyday (fictional) scenarios that involved descriptions of moral transgressions (e.g., cheating on one's partner). Specifically, participants read the scenario context followed by the RSVP of the target sentence containing the critical word (cf. Table 1). The context determined whether the target sentence described a moral or an

Table 1 Example for moral materials with context for moral and for immoral items, for emotional materials with context for neutral and for negative items, as well as the respective target sentences containing the critical word (in italics)

Morality		Emotion	
Moral	Immoral	Neutral	Negative
Herr Zimmermann arbeitet in einem Chemiekonzern. Seit einigen Wochen gibt es Sicherheitsprobleme aufgrund fahrlässigen Verhaltens von Kollegen, sodass Menschenleben gefährdet sind. (Mr. Zimmermann works at a chemical company. For several weeks, there have been safety issues caused by colleagues' careless behavior, putting human lives at risk.)	Herr Zimmermann arbeitet in einem Chemiekonzern als Sachbearbeiter in der Personalabteilung. Er ist ehrgeizig und will rascher aufsteigen als seine Kollegen. (Mr. Zimmermann works in human resources of a chemical company. He is ambitious and aims at being promoted faster than his colleagues.)	Sarah trainierte den ganzen Nachmittag im Fitness-Studio und geht zum Abschluss in die Sauna, wonach sie sich wohlig gut fühlt. (Sarah had been at the gym all afternoon before completing her workout with a visit to the sauna, after which she feels pleasantly well.)	Sarah hat einen sehr anstrengenden Job. Sie muss dafür seit mehreren Monaten jeden Tag um 5 Uhr aufstehen und den ganzen Tag hart arbeiten. (Sarah has a very demanding job. For several months, she has been getting up at 5 a.m. and working very hard all day.)
Im monatlichen Gespräch berichtet er seinem Chef deren <i>Fehler</i> . (In a monthly commitment talk, he reports to his boss their <i>mistakes</i> .) ^a		Sie ist sehr <i>erschöpft</i> . (She is very <i>exhausted</i> .) ^a	
Ben ist wegen sexueller Belästigung angezeigt worden. Er war an dem fraglichen Nachmittag bei einem Freund. Dieser Freund wird von der Polizei befragt. (Ben was sued for sexual harassment. The day in question he was at a friend's. This friend is questioned by the police.)	Ben ist wegen sexueller Belästigung angezeigt worden. Er behauptet an diesem Nachmittag bei einem Freund gewesen zu sein, was jedoch gelogen ist. Sein Freund wird nun befragt. (Ben was sued for sexual harassment. He claims that he was at a friend's, which is a lie. This friend is questioned by the police.)	Dominik fliegt mit seinen Freunden im Hochsommer nach Südspanien. Sie gehen lieber an den Strand als die Stadt zu besichtigen. (Dominik travels by air with his girlfriend in midsummer to the south of Spain. They prefer to go to the beach rather than to visit the town.)	Dominik beeilt sich das Essen für seine wartenden Gäste zu servieren. Ohne zu überlegen greift er in den Kochtopf mit den Eiern, die noch im heißen Wasser schwimmen. (Dominik hurries up to serve food to his waiting guests. Without thinking, he grasps in the cooking pot the eggs, which are still swimming in the hot water.)
Er verschafft ihm das <i>Alibi</i> . (He provided him with the <i>alibi</i> .) ^a		Es ist zu heiß. (It is too <i>hot</i> .) ^a	
Erik studiert Mathematik. Für die Übungsaufgaben hat er sich die Lösungen einer besonders kniffligen Aufgabe beim Professor besorgt. Dieser hat ihn darum gebeten, aus Fairness die Lösungen an die anderen Studierenden weiterzureichen. (Erik majors in math. He asked the professor for the answers to a particularly tricky homework problem. As a matter of fairness, the professor asked him to pass on the answers to his fellow students as well.)	Erik studiert Mathematik. Für die bevorstehende Klausur hat er sich bei einem Freund des höheren Semesters die Lösung besorgt. Dieser hat ihn allerdings darum gebeten die Lösungen nicht herumzureichen, weil er sonst exmatrikuliert werden könnte. (Erik majors in math. He got the answers for an upcoming exam from a friend, who is already a senior. The friend asked him not to pass on the answers to anyone else, otherwise he would get in trouble and have to leave the school.)	Herr Meier hat Streit mit seiner Frau und rennt wütend aus dem Haus. Noch voller Zorn, achtet Herr Meier nicht auf die Straße und sieht den heranfahrenden PKW nicht. (Mr. Meier has an argument with his wife and rushes out of the house in anger. Still upset, he does not mind the traffic and overlooks the approaching car.)	Herr Meier reht den Rasen, während sein Sohn dort mit dem ferngesteuerten Spielzeugauto spielt. (Mr. Meier rakes the lawn while his son plays there with a remote-controlled toy car.)
Er hat den anderen die Lösungen <i>gegeben</i> . (He has to them the answers <i>passed on</i> .) ^a		Herr Meier wird von dem Auto <i>angefahren</i> . (Mr. Meier is by the car <i>hit</i> .) ^a	
Frau Bauer möchte zum Abteilungessen gehen und sucht einen Babysitter. Sie erhält die Anfrage einer Frau, von der man weiß, dass sie gegenüber Kindern bereits handgreiflich wurde. (Mrs. Bauer is looking for a babysitter because she wants to attend the company dinner. A lady contacts her, of whom it is known that she beat children in the past.)	Frau Bauer möchte zum Abteilungessen gehen und sucht einen Babysitter. Sie erhält die Anfrage einer jungen Studentin, die am Telefon sehr sympathisch wirkt, aber mit ausländischem Akzent spricht. (Mrs. Bauer is looking for a babysitter because she wants to attend the company dinner. A young student girl contacts her who sounds very	Magda überrascht ihren Freund, da sie soeben eine Last-Minute-Reise gewonnen hat und diese gerne mit ihm verbringen möchte. (Magda has won a last-minute trip and surprises her boyfriend by inviting him to join her.)	Magda hat wiederholt einen heftigen Streit mit ihrem Freund, der bereits beim letzten Mal drohte sie zu verlassen und aus der Wohnung auszuziehen. Der Disput eskaliert. (Magda and her boyfriend have a serious fight again. The last time he already threatened to leave her and move out. The fight went out of control.)

Table 1 (continued)

Morality		Emotion	
Moral	Immoral	Neutral	Negative
	friendly but speaks with a foreign accent.)		
Sie hat der Babysitterin <i>abgesagt</i> . (She has the babysitter cancelled.) ^a		Er stürmt ins Zimmer und packt den <i>Koffer</i> . (He rushes into the room and packs his <i>suitcase</i> .) ^a	

^a Note that target sentences are translated word-by-word to indicate the position of the critical word, hence disregarding the appropriate word order in English

immoral event. As a control, participants read materials in which the target sentence was either consistent or inconsistent with their knowledge of the world, to assess the ERP correlates elicited by the linguistic processing of moral-neutral world-knowledge violations (e.g., a target sentence of “She receives as a dish a plate full of *snails* and white bread,” following a context that would make this statement either consistent with the participants’ knowledge of the world—e.g., “During a France exchange, Mrs. Lehmann eats a famous French specialty,” or inconsistent—e.g., “Mrs. Lehmann goes to a Schwabian restaurant and orders a local specialty”). Morality and world-knowledge materials were randomly interleaved, and no explicit judgments were required.

Crucially, a larger P200 amplitude was found both for moral transgressions and for world-knowledge violations, indicating domain-unspecific, enhanced attentive processing of materials conflicting with the discourse context. Subsequently, a large posterior N400 was found for general world-knowledge violations only. In accord with previous studies from our lab (e.g., Filik & Leuthold, 2008, 2013) and with the N400 literature in general (cf. Kutas & Federmeier, 2011), this was taken to reflect the increased semantic memory demands involved in retrieving and integrating conceptual information during meaning construction when knowledge-based expectations are violated (e.g., Filik & Leuthold, 2008; Hagoort et al., 2004). By contrast, moral transgressions did not trigger a larger N400 but only a larger central-maximal ERP positivity after about 320 ms. Leuthold and colleagues took this finding to reflect an LPP effect, proposing that incoming socio-normative information is, during a first step, implicitly evaluated and categorized as good or bad (see also Cunningham & Zelazo, 2007). This is in line with theoretical views that assume a central role of emotional-intuitive processes for moral judgment (Greene et al., 2001; Haidt, 2001).

More generally, the ERP study of Leuthold et al. (2015) demonstrates the practicality of approaching the (implicit) mechanisms contributing to moral cognition by having participants read fictional scenarios with moral content. In contrast to Van Berkum et al. (2009), a passive reading task was used in which the moral versus immoral nature of the (identical) target sentences had to be inferred, depending on the discourse context.

That is, the materials did not involve incongruent moral statements but instead described scenarios that participants in a pretest had judged as either clearly morally good versus bad, which would explain the absence of an N400 (congruity) effect. Also, because no explicit moral judgments were required, we consider it more likely that the LPP effect reported by Leuthold et al. reflects the implicit (affective) evaluation of morality-related materials. Emotion effects on the ERP waveform are known to depend on the emotional features of the critical item (e.g., valence, arousal), with emotional stimuli such as positive or negative words, pleasant and unpleasant pictures, and arousing stimuli reliably eliciting larger LPP amplitudes than neutral or less arousing stimuli, and this effect is more pronounced when participants judge the emotional content (e.g., in an affective judgment task) rather than an emotion-irrelevant stimulus dimension (e.g., in a semantic classification or passive reading task; for reviews, see Citron, 2012; Fischler & Bradley, 2006; Hajcak, Weinberg, MacNamara, & Foti, 2012). Hence, it seems reasonable to assume that the LPP elicited in the Leuthold et al. study reflects an emotion effect.

As stated above, the study conducted by Leuthold et al. (2015) did not involve any explicit judgment task. However, there is behavioral evidence suggesting that task demands influence whether an affective versus semantic-cognitive analysis is prioritized (e.g., Lai et al., 2012). Importantly, for the present purposes, evidence from functional magnetic resonance imaging (fMRI) studies corroborates this conjecture for the processing of moral content. That is, fMRI studies consistently indicate that brain areas concerned with both cognitive and emotional processing are activated during moral judgment tasks using dilemma scenarios (e.g., Greene et al., 2001) and socio-normative scenarios (e.g., Moll, de Oliveira-Souza, Bramati, & Grafman, 2002). Crucially, in a meta-analysis of a total of 40 fMRI studies (Sevinc & Spreng, 2014), brain areas concerned with cognitive processing were more strongly activated than areas linked to emotional processing in studies using explicit moral judgment tasks, whereas the reverse pattern of brain activation was found in studies using implicit (e.g., reading) tasks. In line with these findings,

evidence from social cognition research suggests that the impact of automatic evaluations is reduced when participants deliberately rather than implicitly process incoming information (e.g., Bargh, Chaiken, Raymond, & Hymes, 1996). Of course, given the limited temporal resolution of fMRI, the precise time course of task-dependent emotional versus cognitive influences on moral judgments is not yet completely understood (Avramova & Inbar, 2013). Thus, it is essential to investigate this issue using a combined behavioral and ERP approach in order to test whether an explicit moral judgment task would enforce a semantic-cognitive analysis of morality materials, as indicated by the N400. We will address this issue by conducting an experiment in which participants read the materials used by Leuthold et al., but in the context of an explicit morality judgment task.

If one assumes that moral acceptability is inferred from the context and involves the affective evaluation of linguistic input, it is important to assess whether similarly constructed emotional materials without moral content also elicit an LPP effect. At present, we are not aware of any published ERP studies investigating the processing of materials where target sentences are identical across conditions and the emotional meaning of the target needs to be inferred from the context in which it appears. Specifically, previous ERP studies examining discourse-based emotion effects used contexts that were strongly constraining (e.g., León et al., 2010; Leuthold et al., 2012) or employed materials for which the critical words differed across emotion conditions (e.g., Delaney-Busch & Kuperberg, 2013; Holt et al., 2009; León et al., 2010). For example, Delaney-Busch and Kuperberg (2013) found a larger N400 (300–500 ms) for incongruent than for congruent neutral words following a neutral context, and this congruity effect was larger over anterior than over posterior midline electrodes. Crucially, when an emotional discourse context preceded valence congruent or incongruent emotion words, no congruity effect was observed in N400 amplitude. Rather, a larger LPP (500–700 ms) to pleasant and unpleasant emotion words was elicited, irrespective of the valence of the preceding emotional discourse context. In accordance with the affective primacy hypothesis (Storbeck & Clore, 2007), these findings led Delaney-Busch and Kuperberg to suggest that for emotional contexts, the processing of incoming information is dominated by the analysis of their motivational (e.g., approach vs. avoidance) rather than semantic significance. It is therefore unclear whether emotion materials for which the target sentences and the critical (emotion) words are identical across conditions, and hence the emotional meaning has to be inferred from the discourse context, elicit an LPP effect as well. Thus, it is also a major aim of the current work to close this research gap concerning our understanding of emotional language processing, and this is suited to strengthen the interpretation of the LPP as reflecting the affective evaluation of linguistic input during discourse comprehension.

Objectives of the present study

In summary, it remains to be investigated, first, whether the rapid affective evaluation of descriptions of moral transgressions during text comprehension (i.e., when there is no explicit judgment task), as inferred from the LPP effect by Leuthold et al. (2015), is also observed when participants perform explicit moral judgments. If such an LPP effect but no N400 effect would be present, this outcome would lend support to the view that incoming linguistic information undergoes an implicit (or task-independent) affective evaluation. By contrast, if these task conditions enforce a semantic-cognitive analysis of morality materials, a larger N400 to immoral than to moral items should be triggered. Second, it is crucial to investigate the electrophysiological correlates of discourse-based emotion comprehension, specifically, whether an LPP effect is also elicited when the emotional meaning is inferred from the discourse context. The assumption that the LPP indicates the discourse-dependent affective processing of linguistic input, as assumed in the moral ERP study of Leuthold et al. (2015), would be corroborated by showing that for the same participants, discourse-dependent negative compared to neutral (or positive) items elicit a similar LPP effect to discourse-dependent immoral compared to moral items. Therefore, we created novel emotion materials that were similar to the morality materials with regard to critical dimensions, such as the cloze probability of the critical words, their semantic relatedness to the discourse context, critical word frequency, as well as their emotionality in terms of valence and arousal (cf. Method). Of course, because morality and emotion materials differ with regard to the wording of the critical sentences and hence are not matched regarding all potentially relevant word-level or discourse-level dimensions, this allows only an indirect comparison of the ERP effects triggered by these materials.

We recorded ERPs in two experiments to investigate task-related influences on the online processing of scenarios describing everyday moral compared to emotional situations and the nature of the underlying, potentially affective, processes. Thus, our setup was identical to that of Leuthold et al. (2015) except that (a) instead of world-knowledge violations, we used emotional scenarios without moral content as a control condition, and (b) that participants performed explicit judgments of the materials. More specifically, we used prototypical scenarios for which the protagonists and situations were introduced by the context sentences (for an example, see Table 1).

For morality materials, the target sentence either described a morally acceptable or unacceptable (that is, moral vs. immoral) action, and for emotion materials, the target sentence described either a relatively neutral versus a negative event, which was determined by the context for both materials.

We used RSVP for the final critical sentence, with participants performing their judgment response (yes/no) after the presentation of the final word. We chose a binary judgment task in line with recent moral dilemma and moral judgment studies (e.g., Greene et al., 2001). In Experiment 1, participants made moral judgments for morality materials (i.e., “Is the behavior morally acceptable?”) and emotional judgments for emotion materials (i.e., “Are you emotionally moved by the text?”). Experiment 2 required emotional judgments for both types of material.

Generally, we hypothesized that linguistic input is affectively evaluated, that is, independently of the specific content of the materials (cf. Bargh et al., 1996; Cunningham & Zelazo, 2007). It is then reasonable to assume that moral actions and neutral (or mildly positive) events are evaluated as potentially “good” and immoral and negative events as potentially “bad.” Since such differential affective evaluations are taken to be reflected by the LPP component (cf. Fischler & Bradley, 2006; Hajcak et al., 2012), LPP amplitude should be larger for both immoral and negative compared to moral and neutral (or mildly positive) scenarios. We further reasoned that if these evaluations are automatic in the sense that they are produced by a fast-operating process that is independent from task goals, qualitatively the same LPP effects should be observed in the two experiments for immoral versus moral scenarios. However, if the requirement to judge the moral content prioritizes the semantic-cognitive processing of linguistic input in Experiment 1, then a larger N400 rather than a larger LPP might be elicited for immoral than moral items.

Experiment 1

Participants were presented with morality and emotional scenarios in separate blocks of trials. In the case of a morality scenario, they judged whether someone’s behavior was acceptable or not. Here, we predicted that immoral compared to moral scenarios would be judged as less acceptable, hence producing fewer “yes” responses. For emotion materials, participants judged whether they were emotionally moved by the text or not. We predicted that negative compared to neutral scenarios are more moving and therefore would produce more “yes” responses.

It is important to note that we performed rating studies (see Method section for details) to assess the moral acceptability of morality items as well their plausibility, valence, and arousal value and that the same dimensions were assessed for emotion materials (except their moral acceptability). Based on these results, the morality items were classified as either moral or immoral whereas the emotion items were classified as either neutral or negative. This procedure guarantees that the respective materials are neatly matched across conditions. To examine whether item-specific arousal and valence characteristics,

as obtained from the rating studies, contribute to present binary emotion and morality judgments in addition to condition-specific effects, a logistic regression approach was used.

Moreover, because binary morality and emotion judgments (yes vs. no response) are required in the present experiments, it is conceivable that participants may apply decision criteria that lead, at least sometimes, to judgments that are inconsistent with the rating-based morality or emotion classification of the materials. That is, some items preclassified as moral might be judged as morally unacceptable, or some items preclassified as negative might be judged as neutral, and vice versa. Therefore, we performed additional ERP amplitude analyses in waveforms averaged for moral items that were judged as appropriate (“yes” response) and for immoral items judged as inappropriate (“no” response). Likewise, such judgment-dependent ERP analyses were also conducted for neutral items that were judged as not moving (“no” response) and negative items judged as moving (“yes” response).

Method

Participants Thirty-two native German speakers from the University of Tübingen received course credits or payment for participating. Data from four participants were excluded due to excessive alpha activity. For all analyses, we used the data set from the remaining 28 participants ($M = 24.5$ years, 19 females).

Materials and design Morality materials were taken from and modified.¹ These materials consisted of a total of 160 items, resulting from the combination of 80 identical target sentences, each with two different discourse contexts, thereby creating 80 moral and 80 immoral items. The 160 emotion materials were newly generated and analogously constructed (see Table 1 for examples; the full set is available from the first author). Both morality and emotion materials were pretested (see below).

All scenarios consisted of two parts. The first part consisted of two or three sentences describing the context, and the second part was the target sentence containing the critical word. In order to eliminate possible sentence-level and word-based effects, the same target sentence was used for moral and immoral conditions, and the same held true for neutral and negative emotional conditions (with the context varying across conditions; see Table 1). The critical word was always presented toward the end of the target sentence, most frequently as the sentence-final word (84.4%). Critical words were

¹ The original morality materials were changed to disambiguate the meaning of some items and to consistently present critical words toward the end of the target sentence. As a result, two items were replaced. Fifty-two discourse contexts were slightly shortened or the protagonist’s name changed. Of the 80 target sentences, 15 were shortened, and for eight target sentences, the critical word was replaced.

predominantly verbs describing a certain behavior (e.g., to borrow, to report, to mention, to swap) and nouns (e.g., acceptance, alibi, verdict, tumor).

Morality materials described actions that would be perceived as either moral or immoral, whereas emotion materials would describe a neutral or a negative event. Finally, 40 neutral filler items were constructed that contained no moral or emotional content as well as no inconsistencies, and which were similar in length to the experimental items (e.g., Context sentence: “Herr Krüger hat kein aktuelles Telefonbuch. Er braucht die Nummer seines Hausarztes.“ Target sentence: “Er ruft bei der Auskunft an, um an die Nummer zu gelangen.“ [Context sentence: “Mr. Krüger does not possess an up-to-date phone book. He needs the telephone number of his general practitioner.” Target sentence: “He calls the directory enquiries service to find out the number.”]). Following the presentation of the final word, for the morality blocks, the following question was displayed on the screen: “Ist das Verhalten moralisch akzeptabel?” [“Is the behavior morally acceptable?”], and for the emotion blocks: “Berührt dich das Gelesene?” [“Are you emotionally moved by the text?”].² Participants indicated their response (“Ja” [“Yes”] versus “Nein” [“No”]) by pressing the left or right arrow key on the computer keyboard.

The randomization of items and conditions across participants was performed in the following way. The two different types of scenario (morality vs. emotion) were presented in the first versus second half of the experiment, and their order was counterbalanced across participants. For two consecutive participants, two lists were randomly generated such that each morality scenario appeared across the two lists either in the moral or the immoral condition, and each emotion scenario appeared either in the neutral or the negative condition. That is, the two participants received the same target sentence but with a different context in order to manipulate either the morality condition (moral vs. immoral) or the emotion condition (neutral vs. negative). Thus, for each participant, the 200-item list consisted of 40 moral and 40 immoral items, 40 neutral and 40 negative items, as well as 40 neutral filler items. The fillers were included in order to keep the procedure as similar as possible to the study of Leuthold et al. (2015) and to reduce a potential influence of extended local runs of immoral and negative items on ERPs. For instance, after the description of several immoral behaviors, participants might relax their judgment criteria and view immoral acts as more acceptable, which could potentially reduce the N400 effect (cf. Baetens, Van der Cruyssen, Achtziger, Vendekerckhove, & Van Overwalle, 2011).

² There is no one-to-one translation of the German word *berührt* that would match its emotional connotation, yet, in our view, “emotionally moving” comes close. Note that in contrast to the rating study described below, this term was not explicitly related to the state of arousal.

Pretest of materials For the newly created emotion scenarios, we used a Web-based questionnaire to assess the plausibility, valence, and arousal ratings of the materials. Altogether, we recruited 293 undergraduate students from the University of Tübingen ($M = 23.6$ years, 204 females). The 160 scenarios (80 items each with two conditions neutral/negative) were arranged in four lists, each containing 40 randomly arranged scenarios plus the target sentence; each list was rated by no less than 66 participants. Participants were asked to rate on scales from 1 to 8 (a) how plausible they found the scenario (“Die beschriebene Situation ist . . .” [“The scenario described is . . .”]; 1 = *sehr unrealistisch* [very unrealistic] to 8 = *sehr realistisch* [very realistic]), (b) their “Erregungszustand” [arousal] in terms of how much they were emotionally moved by the scenario (1 = *nicht ergreifend* [not moved at all] to 8 = *stark ergreifend* [strongly moved]), and (c) the valence of the materials (1 = *sehr negativ* [very negative] to 8 = *sehr positiv* [very positive]). Two-tailed *t*-tests (cf. Table 2) showed that neutral and negative items were rated as being equally plausible ($M = 6.01$ vs. 6.14), $t(79) = 0.98$, $p = .33$. Furthermore, negative items were rated as being more negative ($M = 2.39$ vs. 5.07), $t(79) = 20.02$, $p < .001$, and more moving than neutral items ($M = 5.00$ vs. 3.51), $t(79) = 9.70$, $p < .001$.

For the morality materials, pretests for plausibility and morality were carried out using a Web-based questionnaire ($N = 55$ participants). On a scale from 1 (*sehr unmoralisch; sehr unrealistisch* [very immoral; very unrealistic]) to 8 (*sehr moralisch; sehr realistisch* [very moral; very realistic]), moral items were rated as being morally more acceptable than immoral items ($M = 5.99$ vs. 2.52), $t(79) = 25.93$, $p < .001$, and also as being slightly more plausible ($M = 6.21$ vs. 5.15), $t(79) = 3.88$, $p < .001$. Additionally, valence and arousal ratings for the morality materials were collected from a fresh group of participants ($N = 40$). On a scale from 1 (*sehr negativ; nicht ergreifend* [very negative; not emotionally moving]) to 8 (*sehr positiv; ergreifend* [very positive; emotionally moving]), moral items were rated as more positive ($M = 5.44$ vs. 2.60), $t(79) = 16.35$, $p < .001$, and less moving ($M = 3.79$ vs. 4.34), $t(79) = 3.20$, $p < .01$, than were immoral items.

To compare valence, arousal, and plausibility scores across the two sets of materials, these rating scores were separately analyzed using ANOVAs with the factors material and condition. Given the above-reported analysis of condition effects, significant results will only be reported for the main effect of material and the interaction of material and condition. For plausibility, the main effect of material, $F(1, 316) = 15.05$, $p < .001$, and the Material \times Condition interaction were significant, $F(1, 316) = 23.88$, $p < .05$, as both neutral and negative emotion items were more plausible than immoral items, all $ps < .001$, but were not more plausible than moral items, all $ps > .21$. For arousal ratings, the Material \times Condition interaction was significant as well,

Table 2 Characteristics and rating data of morality and emotion materials

	Morality		Emotion	
	Moral (<i>SE</i>)	Immoral (<i>SE</i>)	Neutral (<i>SE</i>)	Negative (<i>SE</i>)
Cloze	43% (4%)	38% (4%)	40% (4%)	47% (4%)
Semantic relatedness	0.46 (0.01)	0.45 (0.01)	0.40 (0.01)	0.41 (0.01)
Word frequency	59.46 (9.72)		58.31 (6.98)	
Word length	8.66 (0.21)		7.51 (0.19)	
Plausibility	6.21 (0.03)	5.15 (0.04)	6.14 (0.11)	6.01 (0.11)
Valence	5.44 (0.13)	2.60 (0.09)	5.07 (0.12)	2.39 (0.07)
Arousal	3.79 (0.13)	4.34 (0.11)	3.51 (0.13)	5.00 (0.12)
Morality	5.99 (0.04)	2.52 (0.03)		

Means and standard errors (in brackets) of the rating data were calculated for each material and condition. Plausibility, valence, and arousal ratings concern the whole scenario including the target sentence, from 1 (*very unrealistic, very negative, not touched, very immoral*) to 8 (*very realistic, very positive, strongly touched, very moral*). Only morality materials were rated for the degree of morality from 1 (*very immoral*) to 8 (*very moral*). Word frequency (per million) and word length concern the critical words; cloze probability (as percentages) and semantic relatedness was calculated for the critical word in relation with a context

$F(1, 316) = 14.34, p < .001$, indicating a stronger condition effect for emotion than morality materials. Finally, for valence ratings, the main effect of material was significant, $F(1, 316) = 7.48, p < .01$, due to a lower valence score for emotion compared to morality items.

Moreover, all materials were analyzed with regard to critical word frequency, cloze probability, and semantic relatedness. For calculating word frequencies, we chose the SUBTLEX-DE corpus (Brysbaert et al., 2011). Two words were not listed in the corpus. The frequencies (per million) of the remaining critical words did not differ between materials (morality vs. emotion: $M = 59.46$ vs. 58.31), $t(314) = 0.10, p = .92$.

To determine cloze probability, participants were presented with both the context and the target sentence without the critical word, which they were asked to fill in. Due to an error, no cloze probability scores were obtained for two moral items and one immoral item. Cloze probability did not reliably differ between the moral ($M = 0.43$) and the immoral condition ($M = 0.38$), $t(155) = 1.04, p = .30$, and also not between the neutral ($M = .40$) and the negative condition ($M = 0.47$), $t(158) = 1.36, p = .18$. There were also no significant differences in cloze probability for the critical words of morality materials ($M = .40$) and of emotion materials ($M = 0.43$), $t(315) = 0.73, p = .46$.

Finally, we calculated semantic relatedness as the cosine similarity between the context and the critical word with the LSAfun package in R and the German dewak100k_lsa corpus as semantic space (Günther, Dudschig, & Kaup, 2015) based on Latent Semantic Analysis (LSA; Landauer, Foltz, & Laham, 1998). Seven critical words of the morality materials and five critical words of the emotion materials were not listed in the corpus. In a separate analysis of semantic relatedness scores for the two sets of materials, there was no significant

difference between moral ($M = 0.46$) versus immoral ($M = 0.45$) sentences, $t(144) = 0.09, p = .93$, and also not between neutral ($M = 0.40$) versus negative ($M = 0.41$) sentences, $t(148) = 0.83, p = .41$. However, the comparison between materials revealed a higher semantic relatedness score for morality than for emotion materials ($M = 0.45$ vs. 0.40), $t(294) = 3.44, p < .001$.

Procedure After electrode application, participants were seated in an electrically shielded booth in front of a 21-in. computer monitor (60 Hz) at a viewing distance of 65 cm (maintained by a chin rest). Experimental materials (context, words) were presented at the center of the screen in white 16-point Helvetica font on a black background using the Psychophysics Toolbox extensions (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) running under MATLAB (Release 2012b, The MathWorks, Inc., Natick, MA, USA), on an Apple Mac Mini (OS 10.7). Participants were instructed to avoid any eye, head, and jaw movements and to maintain fixation at the center of the screen during word-by-word presentation. Furthermore, they were instructed to read the stories attentively, and to perform the respective judgments by pressing the appropriate response key.

For each of the morality and emotion materials, a practice block containing three trials preceded the experimental items that were presented in a total of four blocks of 25 items each. Blocks were separated by a short break that was controlled in its duration by the participant. Participants started a trial block by pressing the space bar. Then, the context was displayed for a minimum duration of 1,500 ms. When participants had read the context sentences, they initiated the word-by-word presentation of the target sentence by pressing the space bar, which started with the presentation of a fixation point for 1,000 ms. Then, each word was displayed centrally for 300 ms, with a

200-ms blank interval between successive word presentations. After the offset of the final word and a blank interval of 1,100 ms, the presentation of the decision screen followed, stating the mapping of judgments (yes–no) to response keys. This mapping was constant within a given participant but counterbalanced across participants.

Electrophysiological measures Electroencephalographic (EEG) activity was recorded continuously without online low-pass filtering from 72 Ag-AgCl electrodes using a BIOSEMI Active-Two DC-amplifier system with a sampling rate of 512 Hz for EEG and electrooculogram (EOG). All EEG/ERP analyses were performed using available MATLAB toolboxes (EEGLAB: Delorme & Makeig, 2004; FieldTrip: Oostenveld, Fries, Maris, & Schoffelen, 2011) and custom MATLAB scripts (for details, see Dudschig, Mackenzie, Strozyk, Kaup, & Leuthold, 2016). The analysis epoch started 200 ms prior to the onset of the critical word and lasted for 1,700 ms. For preprocessing purposes, signals from all EEG channels were off-line recalculated to an average reference and high-pass filtered (Butterworth filter, 0.1 Hz, 12 dB/oct). (Ocular) artifacts were then removed, and EEG data were corrected (for a similar procedure, see Nolan, Whelan, & Reilly, 2010). As in Dudschig et al. (2016), a predefined z -score threshold of ± 3 was used to identify outliers relating to channels, epochs, independent components, and single channels in single epochs. Firstly, epochs containing extreme values in single electrodes (e.g., amplifier blockings, values larger ± 1000 μV in any electrode) were removed, as were trials containing values exceeding ± 75 μV in multiple electrodes that were unrelated to eye movements. Secondly, z -scored variance measures were calculated for all electrodes, and noisy EEG electrodes (z score $> \pm 3$) were removed if their activity was uncorrelated to EOG activity. Thirdly, a spatial independent components analysis (ICA) based on the infomax algorithm (Bell & Sejnowski, 1995) was performed on the “cleaned” EEG data set, and ICA components reflecting ocular activity (blinks and horizontal eye movements) were removed from this data set (M [removed components] = 3.4). Fourthly, previously removed noisy channels ($M = 2.35$, range: 0–5) were interpolated in the ICA-cleaned EEG data set using the average EEG activity of adjacent uncontaminated channels within a specified distance (4 cm, ~ 3 –4 neighbors per electrode) in order to ensure a full electrode array for each participant. The mean number of trials remaining ($M = 36.75$ out of 40; range: 21–40, median = 38.0) per condition was not reliably different across conditions, all $ps > .40$.

Data analysis For artifact-free trials, the signal at each electrode site was averaged separately for each experimental condition, time locked to the onset of the critical word, and low-pass filtered (Butterworth filter, 30 Hz, 36 dB/oct). In addition, all

EEG channels were recalculated to an average mastoid reference as in Leuthold et al. (2015) and aligned to a 200-ms baseline prior to the onset of the critical word. To facilitate comparison across studies, similar to previous moral and emotion comprehension studies, mean ERP amplitudes were determined for the following time ranges: 200 to 250 ms (P200; as in Leuthold et al., 2015; Van Berkum et al., 2009); 300 to 500 ms (N400; as in Delaney-Busch & Kuperberg, 2013; Leuthold et al., 2015); and 500 to 700 ms (LPP, as in Delaney-Busch & Kuperberg, 2013; Holt et al., 2009; and similar to Van Berkum et al., 2009). Since P200 effects are typically larger over anterior midline electrodes (e.g., Bohan et al., 2012; Leuthold et al., 2015), whereas N400 and LPP effects usually show a more pronounced centroparietal distribution (e.g., Delaney-Busch & Kuperberg, 2013; Holt et al., 2009; Leuthold et al., 2015; Van Berkum et al., 2009), midline electrodes were pooled to form an anterior (AFz, Fz, FCz) and a posterior region of interest (ROI; CPz, Pz, POz).

Statistical analyses of reaction times and ERP amplitudes were performed by means of repeated-measures analyses of variance (ANOVA). The analysis of the binary yes–no judgments were analyzed using a logit model as recommended by Jaeger (2008), implemented via the `glmer` function within the `lme4` R package (Bates, Maechler, Bolker, & Walker, 2014). Separate `glmer` model fitting procedures were implemented for morality and emotion materials. The model was specified with fixed effects of condition, valence, and arousal with random intercepts for participants and items (i.e., $\text{answer} \sim \text{condition} + \text{valence} + \text{arousal} + (1|\text{participants}) + (1|\text{items})$). For all statistical analyses, the significance level was set to $\alpha = .05$.

Complementing the standard, condition-dependent ERP analysis, judgment-dependent ERP analyses were conducted as mentioned in the introduction. That is, we measured ERP amplitudes in waveforms averaged for moral items that were judged as acceptable (“yes” response) and for immoral items judged as unacceptable (“no” response) as well as for neutral items that were judged as not moving (“no” response) and negative items judged as moving (“yes” response). It is worth mentioning that a possible limitation of this judgment-dependent analysis relates to the fact that, in contrast to the standard analysis, items might not be perfectly matched (i.e., in terms of contexts and critical words presented) across the respective experimental conditions. The ANOVA performed on these ERP amplitude data will be reported after the standard ERP analysis.

Results

Behavioral measures Separate logistic regression analyses were performed for the emotion and morality materials to determine the impact of condition, valence, and arousal for the respective binary judgments. Moral items were more often judged as acceptable than were immoral items (84.91% vs.

17.77%, $p < .001$), and negative items were judged more often as being emotionally moving than were neutral items (68.03% vs. 34.29%, $p < .001$). For moral materials, there was a significant effect of condition ($\beta = -3.23$, $SE = 0.28$, Wald $Z = -11.53$, $p < .001$) and valence ($\beta = 0.38$, $SE = 0.08$, Wald $Z = 4.69$, $p < .001$). These results suggest that the likelihood of “yes” responses (“acceptable”) decreased for immoral items, and increased for more positively rated items.

For emotion materials, there was a significant effect of condition ($\beta = -0.48$, $SE = 0.23$, Wald $Z = -2.06$, $p < .05$), valence ($\beta = -0.30$, $SE = 0.07$, Wald $Z = -4.19$, $p < .001$), and arousal ($\beta = 1.15$, $SE = 0.07$, Wald $Z = 17.15$, $p < .001$), indicating that the likelihood of “yes” responses (“moving”) decreased for neutral items and for more positively rated items but mainly increased for more arousing items.

The separate ANOVAs performed on reaction time (RT) data yielded faster responses to immoral than moral items (1,370 ms vs. 1,603 ms), $F(1, 27) = 6.43$, $p < .05$, $\eta_p^2 = .19$, for the morality materials. For emotion materials, there was a trend for faster responses to negative than neutral items (1,010 ms vs. 1,127 ms), $F(1, 27) = 3.80$, $p = .06$, $\eta_p^2 = .12$.

Electrophysiological measures: Condition-dependent ERP results ERP waveforms averaged according to the rating-based item classification (as determined by the pretests discussed above) are shown in Fig. 1. For the ERP data, we performed separate ANOVAs for morality and emotion materials on mean ERP amplitudes at midline electrodes with factors *condition* (moral vs. immoral or emotional-neutral vs. negative) and *ant-post* (anterior vs. posterior) ROI.

For both types of materials, analyses of mean ERP amplitudes coincided with an overall main effect of ant-post, indicating an anterior positivity for the early P200 time window (200–250 ms), and a posterior positivity for the subsequent time windows (300–500 ms, 500–700 ms). For the sake of brevity, we refrain from reporting the respective main effects of ant-post, all $F_s(1, 27) < 5.77$, all $p_s < .05$, in the following.

200–250 ms (P200) In this time window, there were no reliable condition effects for either morality materials, all $F_s < 1.07$, $p_s > .31$, or for emotion materials, all $F_s(1, 27) < 2.39$, $p_s > .13$.

300–500 ms (N400) Mean ERP amplitudes for morality materials yielded a Condition \times Ant-Post interaction, $F(1, 27) = 5.51$, $p < .05$, $\eta_p^2 = .17$, reflecting a trend for a more negative-going ERP waveform for immoral than moral items over anterior electrodes (3.43 vs. 4.54 μV), $F(1, 27) = 3.87$, $p = .06$, but not over posterior electrodes (4.93 vs. 4.90 μV), $F(1, 27) = 0.01$, $p = .94$.

For emotion materials, ERP amplitudes were more positive going for negative than for neutral items (6.33 vs. 4.17 μV), $F(1, 27) = 13.30$, $p < .001$, $\eta_p^2 = .33$, but the Condition \times Ant-Post interaction was not significant, $F(1, 27) = 2.84$, $p = .10$.

500–700 ms (LPP) In this time window, the reliable Condition \times Ant-Post interaction for morality materials, $F(1, 27) = 6.46$, $p < .05$, indicated a more negative-going waveform for immoral versus moral items for the anterior ROI (4.97 vs. 6.19 μV), $F(1, 27) = 4.61$, $p < .05$, but not for the posterior ROI (7.20 vs. 7.30 μV), $F = 0.03$, $p = .86$.

Finally, mean ERP amplitudes for emotion materials yielded a significant Condition \times Ant-Post interaction, $F(1, 27) = 5.71$, $p < .05$, $\eta_p^2 = .17$. Further testing indicated a larger positivity for negative than neutral items for the posterior ROI (10.04 vs. 8.11 μV), $F(1, 27) = 7.26$, $p < .05$, but not for the anterior ROI (6.36 vs. 6.12 μV), $F = 0.09$, $p = .77$.

Judgment-dependent ERP results ERP waveforms averaged corresponding to the judgment-dependent analysis are shown in Fig. 2. Again, for both types of materials, the analyses of mean ERP amplitudes showed main effects of ant-post, all $F_s(1, 27) < 6.36$, all $p_s < .05$, indicating an anterior positivity for the early P200 time window, and a posterior positivity for the later time windows.

200–250 ms For morality materials, the ANOVA of mean judgment-dependent ERP amplitudes with variables answer (yes vs. no) and ant-post (anterior vs. posterior) produced no significant effects, all $F_s(1, 27) < 1.53$, $p_s > .22$.

The ANOVA for emotion materials showed a trend for the main effect of answer, $F(1, 27) = 3.91$, $p = .058$, $\eta_p^2 = .13$, but no interaction effect, $F < 0.01$, $p > .97$, due to a more positive-going ERP waveform for “no” responses than for “yes” responses (6.44 vs. 5.19 μV).

300–500 ms For morality materials, the Condition \times Ant-Post interaction was significant, $F(1, 27) = 7.06$, $p < .05$, $\eta_p^2 = .21$, indicating a trend towards a more negative-going ERP waveform for “no” responses than for “yes” responses for the anterior ROI (3.55 vs. 4.72 μV), $F(1, 27) = 3.55$, $p = .07$, but not for the posterior ROI (5.13 vs. 5.48 μV), $F = 0.04$, $p = .84$.

For emotion materials, the ERP positivity was larger for “yes” responses than for “no” responses (7.63 vs. 4.17 μV), $F(1, 27) = 14.30$, $p < .001$, $\eta_p^2 = .35$. The answer effect tended to be stronger over posterior than anterior electrodes as indicated by the trend for the Condition \times Ant-Post interaction, $F(1, 27) = 3.73$, $p = .064$, $\eta_p^2 = .12$ (cf. Fig. 2).

500–700 ms For morality materials, the Condition \times Ant-Post interaction was significant, $F(1, 27) = 8.32$, $p < .01$, $\eta_p^2 = .24$, indicating a trend for a more negative-going ERP waveform for “no” responses than for “yes” responses for the anterior ROI (5.22 vs. 6.53 μV), $F(1, 27) = 3.96$, $p = .057$, but not for the posterior ROI (7.91 vs. 7.71 μV), $F(1, 27) = 0.11$, $p = .73$.

For emotion materials, the Answer \times Ant-Post interaction was significant, $F(1, 27) = 6.85$, $p < .05$, $\eta_p^2 = .20$, due to a reliably larger positivity for “yes” responses than for “no”

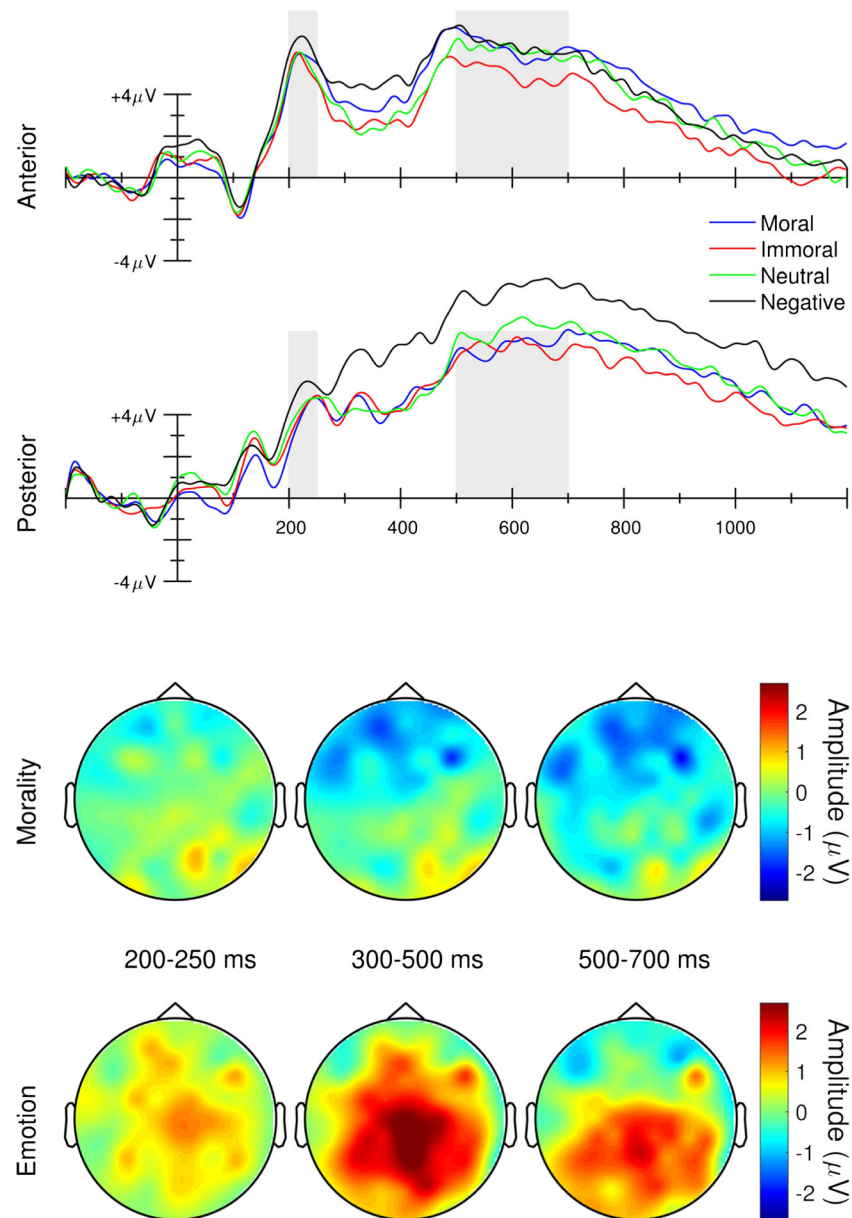


Fig. 1 Upper panel: Condition-dependent grand average ERP waveforms elicited at anterior and posterior midline electrodes, time-locked to the onset of the critical word for morality and emotion materials in Experiment 1. Positivity is plotted upwards. Lower panel: Spline-interpolated topographic

map of mean ERP difference waveform for the 200–250-ms, 300–500-ms, and 500–700-ms time window in Experiment 1. Top row: Morality condition (immoral minus moral). Bottom row: Emotion condition (negative minus neutral). (Color figure online)

responses for the posterior ROI (10.73 vs. 7.61 μV), $F(1, 27) = 9.19$, $p < .01$, but not for the anterior ROI (6.05 vs. 5.10 μV), $F(1, 27) = 1.44$, $p = .24$.

Discussion

In Experiment 1, participants performed different judgments depending on material type, either focusing on the moral acceptability of someone's behavior for morality scenarios or whether they were emotionally moved by the described emotion scenarios. In line with our expectations, behavioral data

showed that moral items were very frequently judged as acceptable and immoral items as unacceptable, with less than 18% of the items being judged by participants in a way that was inconsistent with the classifications that were based on the results of the pretest (e.g., judging an item of the immoral condition as morally acceptable). Crucially, the logistic regression analysis indicated that, in addition to the preclassified condition variable, valence also influences morality judgments, in line with views that emotional aspects of the scenarios contribute to moral decision-making (e.g., Greene et al., 2001; Haidt, 2001). For emotion items, subjective judgments

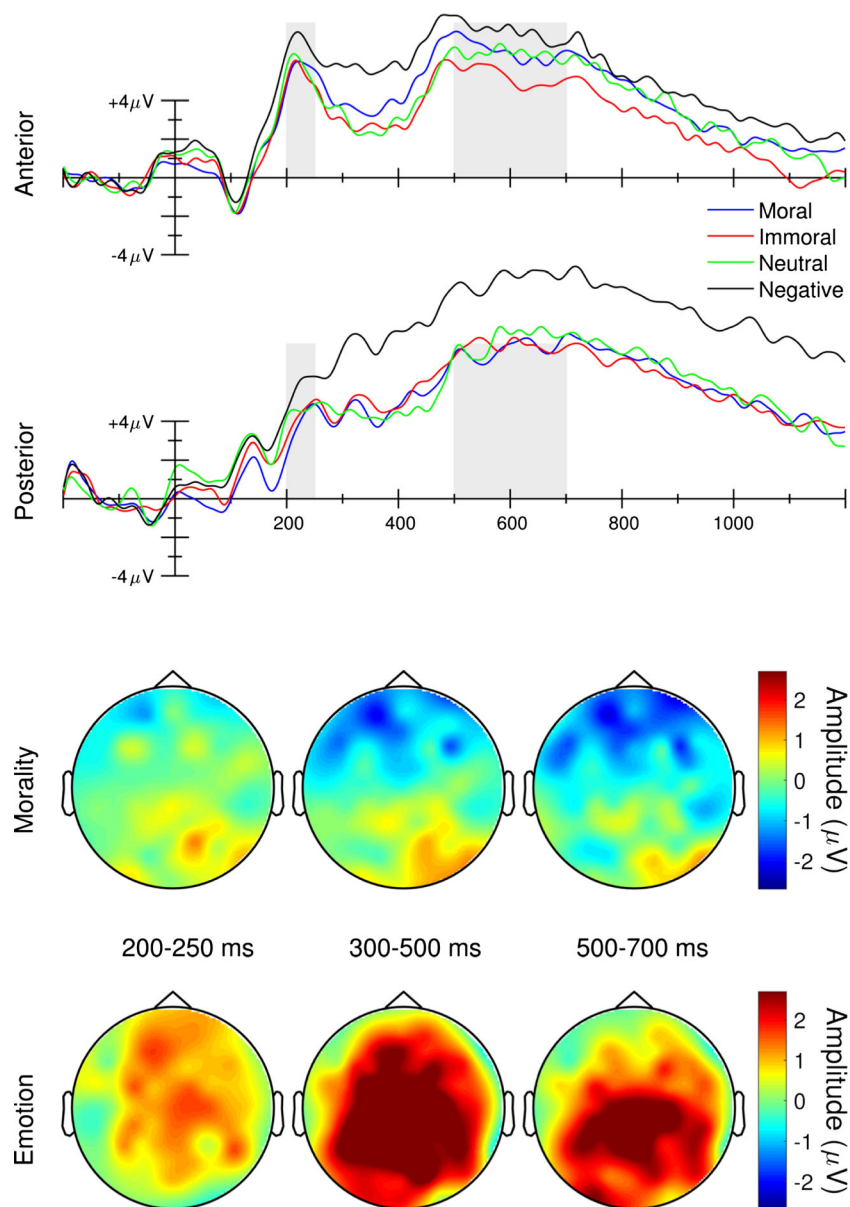


Fig. 2 Upper panel: Judgment-dependent grand average ERP waveforms elicited at anterior and posterior midline electrodes, time-locked to the onset of the critical word for morality and emotion materials in Experiment 1. Positivity is plotted upwards. Lower panel: Spline-interpolated topographic

map of mean ERP difference waveform for the 200–250-ms, 300–500-ms, and 500–700-ms time window in Experiment 1. Top row: Morality condition (yes minus no). Bottom row: Emotion condition (yes minus no). (Color figure online)

of items did accord slightly less well with the preclassified neutral versus negative item classification (about 67%). A possible reason for this lower consistency is suggested by the logistic regression analysis results, which indicated that mainly rating-based arousal scores and to a lesser extent valence scores for each item were influencing the affective yes–no judgments in addition to the preclassified condition variable. This is also plausible, given the rating results for emotion materials, indicating that some neutral scenarios received positive valence ratings. There was also a moderately positive correlation indicating increasing arousal ratings with positive

valence for these neutral items ($r = .32$). Finally, another reason could be that some of the sentence final emotional words were valenced, and thus “yes” responses to neutral items might also reflect a word-based valence effect.

The finding of shorter RTs for immoral and negative items indicates that these items were more salient than moral and neutral items, as also suggested by the rating study results. In addition, we observed faster responses for emotional than for moral materials, suggesting that moral judgments involve a more time-consuming decision process. However, based on this result alone, we cannot exclude the possibility that

affect-related processes contribute to moral judgments. In summary, the behavioral data clearly indicate that participants performed the different judgment tasks appropriately.

In terms of the ERP results, a first key finding relates to the ERP analysis for emotion materials, which showed a larger posterior than anterior ERP positivity from 300 ms to 700 ms, as expected. Given its topographic distribution and time course, and the fact that its amplitude was larger for negative than for neutral materials, we take this positivity to reflect the LPP. It is worth noting that the judgment-dependent analysis of ERP amplitudes produced the same results as the standard analysis of ERP amplitudes. Hence, we take the larger LPP to negative than neutral items to reflect an emotion effect. This inference seems justified given the pretest results for emotion materials. That is, negative and neutral emotion items differed with respect to their valence and arousal but not regarding their linguistic features (cloze probability, semantic relatedness, critical word frequency). Similar to Van Berkum et al. (2009), we view it as unlikely that the present LPP emotion effect reflects a decision-related P300 effect; we return to this issue in the General Discussion. Together, and in line with similar reports in the literature (e.g., Delaney-Busch & Kuperberg, 2013; Holt et al., 2009), we therefore take the present LPP findings as reflecting stronger affective processing of negative than neutral items during discourse comprehension.

Crucially, and in contrast to our hypothesis that morality items undergo an implicit affective evaluation as indicated by an LPP effect (cf. Leuthold et al., 2015), no reliably enhanced ERP positivity was observed for immoral as compared to moral items. Rather, a more negative-going ERP amplitude for immoral than for moral scenarios appeared from 500ms to 700 ms (cf. Fig. 1). Although the direction of this amplitude effect is in line with the centroparietal N400 effect reported by Van Berkum et al. (2009), its topographic distribution is not. That is, the present morality effect in the 500–700 ms time window showed an anterior rather than the classic centroparietal N400 distribution, it occurred later, and was also more sustained. In the General Discussion, we will evaluate possible explanations for this anterior ERP negativity effect.

In summary, the ERP findings from Experiment 1 indicate that evaluative processing of immoral items elicited a larger anterior ERP negativity than moral items, whereas negative emotional items triggered a larger LPP than neutral items, suggesting that the different materials differ with regard to their cognitive versus affective processing. This difference in processing might be attributed to the fact that participants performed different tasks to the two types of materials. In the following experiment, we will therefore test whether the evaluation of moral content, as indicated by the anterior negativity, is task dependent by asking participants to perform emotional judgments for morality materials as well.

Experiment 2

In Experiment 2, participants saw the same morality and emotional scenarios as in Experiment 1, but in this case judged all materials as to whether they were emotionally moved by them. We reasoned that focusing on the evaluation of the emotional content of morality materials might change their online processing in such a way that their affective analysis was prioritized (e.g., Lai et al., 2012; see also Holt et al., 2009). Like in Experiment 1, we analyzed the binary yes–no judgments using logistic regression analyses. This also allowed us to examine whether task demands influence the processing of morality materials and whether emotion judgments for both morality and emotion materials take into account the same affective item dimensions. In this case, a larger LPP should be elicited by both immoral and negative emotional items compared to moral and neutral emotional items. That is, the larger anterior negativity observed for morality materials in Experiment 1 should be absent.

Method

Participants Thirty right-handed native German speakers from the University of Tübingen participated for course credits or payment. Data from one participant were excluded from the analyses due to less than 30% of trials per condition remaining after artifact rejection and from one participant due to excessive alpha activity. Because one behavioral data set was lost due to a technical problem, 27 participants entered the behavioral data analysis, and 28 participants ($M = 23.0$ years, 20 females) contributed data to the ERP analysis.

Materials, procedure, and design Experiment 1 was identical to Experiment 2 concerning all methodological aspects, except that participants now performed yes–no responses to both moral and emotion materials with regard to the question: “Berührt Sie das Gelesene?” [“Are you emotionally moved by the presented text?”].

Data analysis During EEG preprocessing, the number of ICA components removed for cleaning the EEG data set was $M = 3.9$, and the number of previously removed noisy channels that were interpolated in the ICA-cleaned EEG data set was $M = 1.7$ (range: 0–5). Following artifact rejection, the mean number of trials remaining per condition ($M = 37.50$ trials out of 40; range: 27–40, median = 39.0) was not reliably different across conditions, all $ps > .29$.

Binary yes–no judgments were analyzed using a logistic regression analysis identical to that of Experiment 1. Also, in addition to the standard ERP analysis, we measured ERP amplitudes in waveforms averaged for moral and neutral items that were judged as not emotionally moving and for immoral and negative items judged as moving. These judgment-

dependent ERP results are reported at the end of the results section.

Results

Behavioral measures As in Experiment 1, separate logistic regression analyses were performed for the binary emotion judgments to emotion and morality materials. For morality materials, immoral items were more often judged as emotionally moving than were moral items (71.85% vs. 44.30%, $p < .001$). There were significant effects of condition ($\beta = 0.83$, $SE = 0.22$, Wald $Z = 3.71$, $p < .001$), valence ($\beta = -0.15$, $SE = 0.06$, Wald $Z = -2.28$, $p < .05$), and arousal ($\beta = 0.94$, $SE = 0.08$, Wald $Z = 11.13$, $p < .001$). These results suggest that the likelihood of “yes” responses (“moving”) increased for immoral items and for more arousing items but slightly decreased for more positively rated items.

For emotion materials, negative items were more frequently judged as moving compared to neutral items (72.69% vs. 35.13%, $p < .001$). There was a significant effect of valence ($\beta = -0.18$, $SE = 0.08$, Wald $Z = -2.17$, $p < .05$) and arousal ($\beta = 1.34$, $SE = 0.08$, Wald $Z = 16.77$, $p < .001$), indicating that the likelihood of “yes” responses (“moving”) slightly decreased for more positively rated items but mainly increased for more arousing items.

The ANOVA performed on RT yielded no reliably faster responses to immoral than to moral items (1,044 ms vs. 1,157 ms), $F(1, 26) = 2.82$, $p = .11$, $\eta_p^2 = .10$. For emotion materials, RT was faster for negative than for neutral items (953 ms vs. 1,054 ms), $F(1, 26) = 6.15$, $p < .05$, $\eta_p^2 = .19$.

Electrophysiological measures: Condition-dependent ERP results ERP waveforms averaged according to the rating-based item classification are shown in Fig. 3. For both types of materials, the waveform was characterized by an anterior P200 (200–250 ms), and a broadly distributed positivity between 300–500 ms that tended to be posteriorly distributed in the late LPP time window (500–700 ms). As before, main effects of topography will not be discussed in the following.

200–250 ms (P200) Mean ERP amplitudes for morality materials in this time interval were not reliably influenced by experimental conditions, all $F_s < 0.31$, $p_s > .58$.

For emotion materials, the Condition \times Ant-Post interaction was significant, $F(1, 27) = 4.81$, $p < .05$, $\eta_p^2 = .15$, indicating a larger positivity for negative than for neutral items for the posterior ROI (4.32 vs. 2.75 μV), $F(1, 27) = 13.87$, $p < .001$, and as a trend for the anterior ROI (7.03 vs. 6.28 μV), $F(1, 27) = 3.33$, $p = .08$.

300–500 ms (N400) Mean ERP amplitudes for morality materials yielded a Condition \times Ant-Post interaction, $F(1, 27) = 5.88$, $p < .05$, $\eta_p^2 = .18$. Further testing indicated a trend for a

more positive-going ERP waveform for immoral than for moral items for posterior ROIs (5.63 vs. 4.91 μV), $F(1, 27) = 3.73$, $p = .06$, but no reliable effect for anterior ROIs (4.85 vs. 4.46 μV), $F < 0.01$, $p = .98$.

For emotion materials, the significant Condition \times Ant-Post interaction, $F(1, 27) = 20.17$, $p < .001$, $\eta_p^2 = .43$, indicated that the condition effect was more pronounced for the posterior ROI. However, the ERP positivity was reliably larger for negative than for neutral items over both anterior (6.69 vs. 5.67 μV) and posterior midline electrodes (7.84 vs. 4.88 μV), all $F_s(1, 27) > 6.57$, $p < .05$.

500–700 ms (LPP) For the morality materials, in this time window only the interaction of Condition \times Ant-Post was significant, $F(1, 27) = 5.72$, $p < .05$, $\eta_p^2 = .17$, but further testing revealed no reliable effects, all $F_s(1, 27) < 2.93$, $p_s \geq .10$.

Finally, analyses of emotion materials showed a significant Condition \times Ant-Post interaction, $F(1, 27) = 26.32$, $p < .001$, $\eta_p^2 = .49$, due to a larger positivity for negative than neutral items for the posterior ROI (10.09 vs. 7.57 μV), $F(1, 27) = 35.50$, $p < .001$, but not for the anterior ROI (7.98 vs. 7.80 μV), $F(1, 27) = 0.20$, $p = .66$.

Judgment-dependent ERP results ERP waveforms averaged according to the judgment-dependent classification are shown in Fig. 4.

200–250 ms The ANOVA for judgment-dependent ERPs of morality materials revealed no reliable condition effects, all $F_s(1, 27) < 1.69$, $p_s > .20$.

For emotion materials, there was a trend for a larger positivity for “yes” responses versus “no” responses (5.25 vs. 4.43 μV), $F(1, 27) = 3.09$, $p = .09$.

300–500 ms In this time window, ERP amplitudes of morality materials yielded a larger positivity for “yes” responses than for “no” responses (5.68 vs. 4.45 μV), $F(1, 27) = 4.66$, $p < .05$, $\eta_p^2 = .15$.

For emotion materials, the Answer \times Ant-Post interaction was significant, $F(1, 27) = 11.79$, $p < .01$, $\eta_p^2 = .30$, due to an enlarged positivity for “yes” responses than for “no” responses for the posterior ROI (7.65 vs. 4.27 μV), $F(1, 27) = 38.38$, $p < .001$, compared to the anterior ROI (6.33 vs. 4.91 μV), $F(1, 27) = 5.16$, $p < .05$.

500–700 ms ERP amplitudes in this subsequent time window were not influenced by experimental condition for morality materials, all $F_s(1, 27) < 2.35$, $p_s > .13$.

For emotion materials, the Answer \times Ant-Post interaction was significant, $F(1, 27) = 13.75$, $p < .001$, $\eta_p^2 = .34$; further testing an enlarged positivity for “yes” responses than for “no” responses for the posterior ROI (9.74 vs. 7.49 μV),

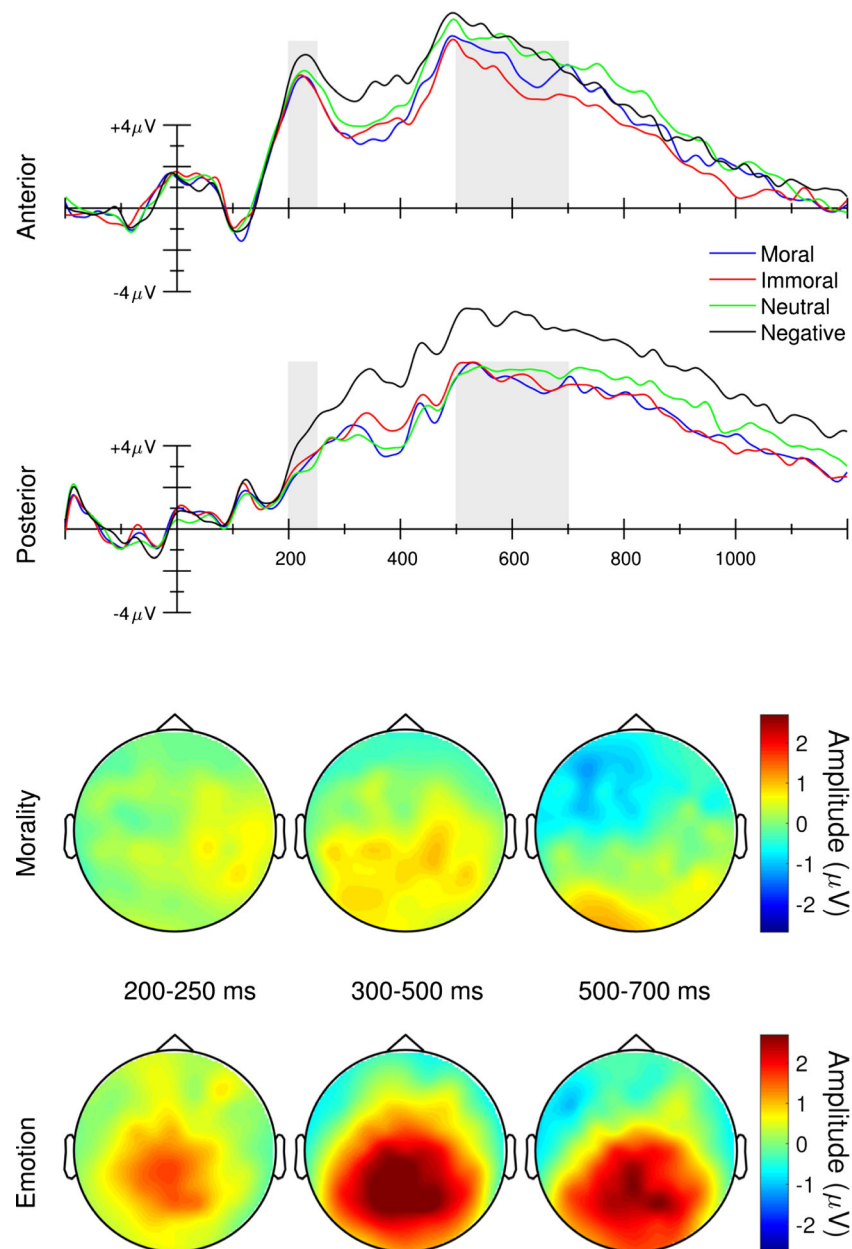


Fig. 3 Upper panel: Condition-dependent grand average ERP waveforms elicited at anterior and posterior midline electrodes, time-locked to the onset of the critical word for morality and emotion materials in Experiment 2. Positivity is plotted upwards. Lower panel: Spline-

interpolated topographic map of mean ERP difference waveform for the 200–250-ms, 300–500-ms, and 500–700-ms time window in Experiment 2. Top row: Morality condition (immoral minus moral). Bottom row: Emotion condition (negative minus neutral). (Color figure online)

$F(1, 27) = 7.75, p < .01$, but not for the anterior ROI (7.66 vs. 7.53 μV), $F = 0.04, p = .85$.

Discussion

The behavioral data analysis showed that as in Experiment 1, responses were faster for negative than for neutral items, again lending support to the conclusion that the former items are emotionally more salient. In addition, the RT analysis indicated that moral and immoral items did not reliably differ with regard to the speed of emotion judgments, whereas they did

for morality judgments in Experiment 1. Importantly, response behavior differed with regard to the preclassified item category for morality and emotion materials. Therefore, as in Experiment 1, ERP amplitudes were analyzed dependent on the preclassified item categories (standard analysis) and dependent on the actual judgments.

First, however, it is important to note that the logistic regression analysis results for morality materials indicated that affective judgments were influenced not only by condition but mainly by rating-based arousal scores and to a smaller extent by valence scores for each item. Thus, in conjunction with the

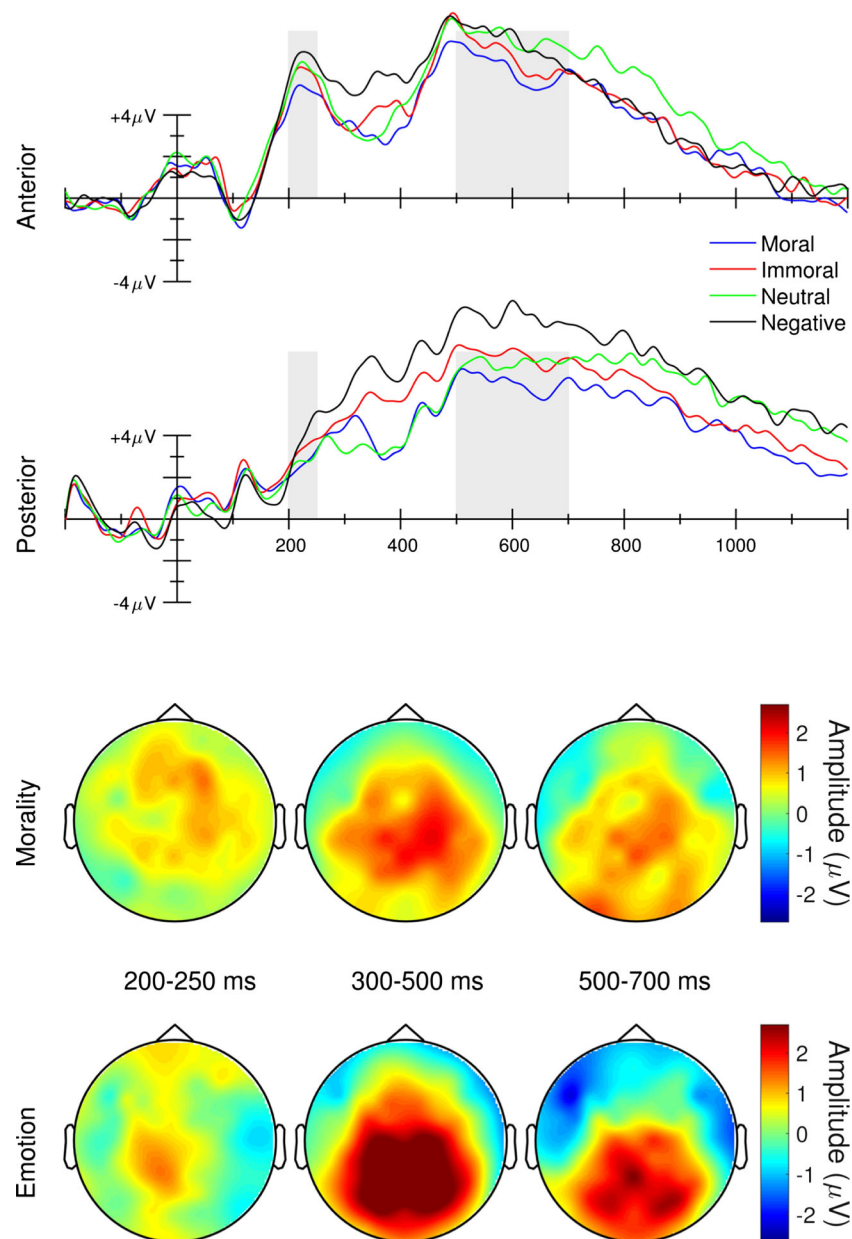


Fig. 4 Upper panel: Judgment-dependent grand average ERP waveforms elicited at anterior and posterior midline electrodes, time-locked to the onset of the critical word for morality and emotion materials in Experiment 2. Positivity is plotted upwards. Lower panel: Spline-

interpolated topographic map of mean ERP difference waveform for the 200–250-ms, 300–500-ms, and 500–700-ms time window in Experiment 2. Top row: Morality condition (yes minus no). Bottom row: Emotion condition (yes minus no). (Color figure online)

logistic regression analyses results for emotion materials, it appears that emotion judgments are strongly influenced by arousal and less so by valence. It is hence understandable that moral as compared to neutral items were judged more frequently as moving (44.30% vs. 35.13%; $p < .001$). This finding is plausible given the fact that arousal rating results for morality materials indicated a smaller difference between moral and immoral items than neutral and negative items, which might also account for the absence of a reliable RT effect for the morality materials.

In the standard ERP analysis, replicating the LPP findings from Experiment 1, there was a larger posterior positivity for negative than for neutral items from 300 ms to 700 ms, which we again take to reflect the LPP. This finding is consistent with the earlier conclusion that the LPP effect reflects the (affective) evaluation of motivationally significant stimuli (e.g., Hajcak et al., 2012). Moreover, rather than an anterior negativity (500–700 ms) as observed in Experiment 1, immoral compared to moral items tended to elicit a larger ERP positivity in the 300 ms to 500 ms time window over posterior midline electrodes. Importantly, corroborating this LPP effect,

the judgment-dependent ERP analysis revealed a larger posterior positivity in the 300–500 ms time window for both immoral and negative items judged as moving as compared to moral and neutral items judged as nonmoving. Thus, LPP findings from Experiment 2 indicate that the affective evaluation of incoming linguistic information occurs not only for emotion materials but also morality materials.

General discussion

In two ERP experiments, we investigated the nature and time course of evaluative processing of short morality (moral vs. immoral) and emotion (neutral vs. negative) scenarios using a discourse comprehension paradigm. Participants judged whether they found the described moral situation as either morally acceptable or not and the emotional situation as moving or not (Experiment 1), or made emotional judgments to both types of scenarios (Experiment 2). Assuming that affective evaluations are triggered by both morality and emotion materials, we expected that critical words would trigger an early enhanced LPP (starting at ~300 ms) for immoral compared to moral scenarios and for negative compared to neutral emotion materials, irrespective of the judgment task. If, however, performing moral judgments (Experiment 1) shifts the focus to the cognitive-semantic processing of moral content, we assumed that an N400 effect might instead be triggered by morality materials.

Crucially, we obtained behavioral evidence for the task-dependent processing of morality materials and also that the specific emotional characteristics (valence, arousal) of emotion and morality materials influenced participants' judgments. Specifically, moral acceptability judgments were slower for moral than for immoral items, whereas the speed of emotional judgments did not reliably differ. The latter judgments were also performed faster than the morality judgments. On the one hand, this outcome suggests that a more complex and hence time-consuming cognitive decision process underlies moral than emotional decision-making, at least for the materials used in this study. On the other hand, it also indicates that the way readers process information about the persons and events described in the text, that is, which information they focus on and evaluate, depends on their specific goals. The additional finding of faster judgment responses to emotionally negative and immoral items than to neutral and moral items in Experiment 1 might be attributed to the fact that the former items are more salient. Finally, in both experiments, participants answered the respective judgment questions as expected, in the majority of cases. That is, immoral items were judged as less acceptable (Experiment 1) and more moving (Experiment 2) than were moral ones, and negative items were judged as emotionally more moving than were neutral ones. Still, binary judgment behavior differed as compared to the

preclassified item category for both morality and for emotion materials. These findings suggest that participants adopted response criteria that did not fully accord with the rating-based classification of items. Whereas the rating study suggested that immoral compared to moral items have higher mean valence and mean arousal scores, it is clear that there is no perfect separation of moral and immoral conditions with regard to these emotion dimensions at the level of individual items, as outlined earlier. Moreover, deciding whether a scenario is morally acceptable or not as well as being either moving or not might involve the processing of stimulus aspects different from those defining their moral content alone. This assumption is supported by logistic regression analysis results. That is, moral judgments were influenced by valence but not arousal, whereas emotion judgments were mainly driven by differences in arousal rather than valence for both morality and emotion materials. Together, behavioral findings indicate that participants followed task instructions and, more importantly, that processing was influenced by the task and the specific moral and emotional content of materials, which is why ERP amplitudes were also analyzed dependent on both the preclassified item categories (standard analysis) and dependent on the actual judgments.

A first key ERP finding concerns the larger LPP for negative than for neutral emotional scenarios, starting after about 300 ms and lasting at least up to 700 ms after the onset of the critical word. It is also important to note that this LPP effect replicated across two independent experiments using the same materials and tasks. Crucially, target sentences were identical for negative and neutral items, and the discourse contexts were only moderately constraining regarding the critical word. Hence, the observed LPP effects reflect a discourse-based influence and are not the result of mere lexical differences between target words or expectancy-driven processes that would be indicated by the N400 or the P300 components. In accord with similar previous research (e.g., Fields & Kuperberg, 2012; Holt et al., 2009; see also Fischler & Bradley, 2006), we therefore take this long-lasting LPP effect to indicate the more intense affective evaluation of negative than of neutral items.

In this respect, the present work extends previous ERP studies examining discourse-based emotion effects using contexts that were either strongly constraining, and hence presumably triggered emotion congruity effects as indicated by the N400 (e.g., León et al., 2010; Leuthold et al., 2012), or varied the critical words across emotion conditions (e.g., Delaney-Busch & Kuperberg, 2013; Fields & Kuperberg, 2016; Holt et al., 2009; León et al., 2010). For instance, Delaney-Busch and Kuperberg found a larger LPP (500–700 ms) to pleasant and unpleasant emotion words, irrespective of the valence of the preceding emotional discourse context, whereas the N400 effect was absent. They interpreted this finding in terms of the affective primacy hypothesis

(Storbeck & Clore, 2007) and proposed that for emotional contexts the affective processing of incoming information dominates over semantic processing. The present LPP effect, in conjunction with the absence of an N400 effect, accords with this view, suggesting that participants focused on the processing of the emotional rather than the semantic content in the present affective judgment task (e.g., Lai et al., 2012). Together, the present ERP findings for emotion materials narrow the identified research gap concerning the investigation of emotional language comprehension by demonstrating that an LPP indicating more intense affecting processing is also observed when discourse contexts determine the emotional meaning of identical critical (emotion) words in target sentences. However, the functional interpretation of the LPP as reflecting the affective processing of linguistic input is still a matter for further research (see below).

Importantly, we reasoned that if evaluative-affective categorization (as indicated by the LPP) contributes to moral judgments, then we should also see a larger LPP for immoral than for moral items, as in previous studies using a similar approach (Leuthold et al., 2015; Van Berkum et al., 2009). In fact, such an enhanced LPP for immoral compared with moral items was present over posterior electrodes from 300 ms to 500 ms in Experiment 2. However, before discussing potential implications of this ERP effect, it is important to consider the alternative possibility that it reflects an N400 effect. In this case, one would have to assume that a larger N400 to moral than to immoral items overlaps with the positive-going ERP waveform, thereby producing a larger ERP positivity to immoral than to moral items. For instance, Holt et al. observed a larger N400 to negative and positive words compared with neutral words, but only if participants passively read for comprehension. When they evaluated the emotional content, however, this N400 amplitude modulation was obscured by the overlapping LPP. This possibility would require that moral as compared with immoral items produce a cost at the level of lexicosemantic processing or when accessing semantic memory, as it is typically the case for incongruent items with low cloze probability or low LSA scores. Yet the present materials used identical target sentences, which precludes the influence of word-based effects. Moreover, an N400 effect due to material differences at sentence and discourse level is not supported, since moral and immoral items did not differ with regard to cloze probability and LSA scores. Also, Leuthold et al. (2015) found an LPP effect and no sign of an N400 when using a passive text comprehension task for which Holt et al. (2009) found an N400 effect to context-incongruent emotion materials. Together, we view it unlikely that the present ERP effect is due to an N400 component overlap and rather reflects a genuine LPP effect. Thus, it appears that participants not only judged immoral items as emotionally more moving than moral ones but also that these items underwent more intense affective processing. We did not find a larger LPP for immoral

than moral items, however, when moral judgments were required in Experiment 1. This finding accords with other discourse comprehension studies in that the LPP, and hence affective processing of linguistic input, is modulated by various variables, including the specific discourse context and task demands (e.g., Delaney-Busch & Kuperberg, 2013; Fields & Kuperberg, 2016; Holt et al., 2009; Xiang & Kuperberg, 2015).

Critically, the moral judgment task had an impact on online processing, as suggested by the ERP findings of Experiment 1, in which an anterior negativity (rather than the LPP), differed in amplitude across morality conditions.³ Before discussing the possible functional significance of this negative ERP deflection in more detail, it is helpful to first rule out a possible alternative explanation in terms of ERP component overlap. Specifically, since N400 and LPP effects are known to be similarly distributed over the scalp, there remains a possibility that simultaneously triggered LPP and N400 effects attenuate each other, with the N400 effect showing up only over anterior electrodes. However, we consider this rather unlikely for the following reasons. First, the present anterior negativity effect was more sustained than typical N400 effects. Second, there were only relatively small LPP effects for morality materials in Experiment 2, despite the fact that emotional judgments were required, which are known to increase LPP effects in comparison to a passive comprehension task (e.g., Holt et al., 2009). Third, cloze probability for critical words and the target sentence (as well as semantic similarity) was the same for moral and immoral items, thereby minimizing possible (predictive) sentence-level and word-based effects on information processing, which are known to trigger a posteriorly distributed N400 effect. Finally, what mattered in our materials were the moral implications of the events being described, whereas the posterior N400 effect in Van Berkum et al.'s (2009) study was triggered by explicit moral statements that were value-incongruent rather than congruent.

We observed that the present immoral compared with moral items elicited a tentatively larger negative-going deflection

³ When analyzing the present judgment-dependent midline ERP amplitudes for moral materials in Experiment 1 and Experiment 2 for the 300–500 ms time window, the ANOVA with the between-subjects factor task (moral judgment vs. reading) and the repeated measurement factors condition (moral vs. immoral) and ant-post (anterior vs. posterior) revealed a significant Task \times Condition interaction, $F(1, 54) = 4.67, p < .05$, which was also significant for the anterior ROI, $F(1, 54) = 4.52, p < .05$. Moreover, analysis of midline ERP amplitudes for moral materials in Experiment 1, in combination with those of Leuthold et al. (2015) for the 300–500 ms time window, the analogous ANOVA with the between-subjects factor task (moral judgment vs. reading) and the repeated-measurement factors condition and ant-post revealed a significant Task \times Condition interaction, $F(1, 54) = 6.47, p < .05$. This was because ERP amplitudes were more negative going for immoral than for moral items in the moral judgment task, but a reverse amplitude effect in the reading task of Leuthold et al. (2015). This is an interesting result because it also indicates that the ERP amplitude effect to moral items depends on the task.

over anterior electrodes from roughly 300 ms to 700 ms after critical word onset. Of course, since this morality effect on the anterior negativity was unexpected, it is important to replicate this ERP effect in future studies and to elaborate its potential functional interpretation. In the following, we present such a possible interpretation based on other discourse comprehension studies that also found an anterior negativity (Baggio, van Lambalgen, & Hagoort, 2008; Xiang & Kuperberg, 2015). In these studies, the anterior negativity was taken to index language-related working memory demands, that is, when alternative but likely text inferences have to be simultaneously maintained or integrated within the discourse or situation model (cf. Zwaan & Radvansky, 1998). Specifically, we therefore speculate that when explicit moral judgments are required, this might impact on the processing of scenarios and the updating of the discourse model in such a way that readers maintain in working memory for a short while after critical word input both the moral and immoral action (for a similar reasoning, see Xiang & Kuperberg, 2015). Put differently, it is possible that working memory load and the demands on integrating linguistic information into the discourse model is higher in the case of immoral than in moral items, giving rise to the enlarged anterior negativity.

Certainly, assuming that the present anterior negativity effect might relate to working memory functions would imply that cognitive-semantic processing plays a role when explicit moral judgments are required. By contrast, when participants merely read the same moral materials for comprehension instead of performing an explicit moral judgment task (Leuthold et al., 2015), a larger LPP was elicited by immoral than by moral items, which we took to reflect the affective evaluation of morality materials.⁴ Thus, it is evident that discrepant ERP patterns result, indicative of cognitive (anterior negativity) and affective processing (LPP), when explicit moral judgments are required rather than when the moral content is implicitly processed. Such a task-dependent impact on moral information processing is in line with fMRI evidence indicating that cognitive processes are more dominant when the task requires explicit moral judgments than merely the passive processing of moral content and vice versa (Sevinc & Spreng, 2014).

Open issues

An open issue concerns the question of whether, and in which way, the LPP is related to the P300 component. For instance, it is known that the amplitude of the centroparietal P300 is inversely related to the prior and also the subjective probability of a given stimulus event, task demands, and the significance

of stimulus input (e.g., Johnson, 1988). With regard to ERP studies using emotional discourse contexts to study person perception (Bartholow, Fabiani, Gratton, & Bettencourt, 2001; Van Duynslaeger, Van Overwalle, & Verstraeten, 2007), it is interesting to note that a larger centroparietal ERP positivity has been found to sentence-final words describing a trait-consistent (“... gave his wife a slap”) than a trait-inconsistent behavior (“... gave his mother a kiss”) following a short passage of text describing a person (e.g., as being hostile). Assuming that readers construct a situation model in working memory about the persons and events described in the text, in line with theories about the mental processes reflected by the P300 (cf. Donchin & Coles, 1988; Nieuwenhuis, Aston-Jones, & Cohen, 2005), one might then assume that a larger P300 (or LPP) is triggered if this model needs updating, as in the case of inconsistent language input. With regard to the impact of emotional stimulus characteristics, the more recent locus coeruleus (LC)-P300 theory (e.g., Nieuwenhuis et al., 2005) might provide an integrative framework for the interpretation of the P300 and the emotion-related LPP, since it assumes that the centroparietal positivity reflects a phasic, LC-mediated enhancement of cortical activity not only after unexpected but also after motivationally relevant and salient stimuli.

It is also an open issue whether the integration of linguistic information into the discourse or situation model is reflected by ERP negativities rather than ERP positivities. Thus, the N400 has also been related to the demands of integrating linguistic input into a situation model (e.g., Nieuwland & Van Berkum, 2006; Filik & Leuthold, 2008, 2013). Moreover, we speculated above that the present anterior negativity might also reflect such integration demands. Together, it remains an important task to further examine the cognitive and affective processes that are more specifically reflected by various ERP components (P300, LPP, N400, and anterior negativity) typically observed in discourse comprehension studies.

Conclusions

In conclusion, the present study provides evidence for the assumption that the processing of morality scenarios depends on the specific task performed by participants. Specifically, for explicit moral judgments, immoral items elicited a larger anterior negativity than did moral items, indicating the enhanced cognitive processing of moral content. By contrast, an LPP effect similar to that observed for negative compared to neutral emotional items was elicited for emotion judgments, indicating the affective categorization of incoming information during discourse comprehension. Future research would need to take into account the potential impact of task demands when elucidating the nature of the potential cognitive and affective processes contributing to moral evaluations and decisions.

⁴ We would like to note that in a current text comprehension study (in preparation), using the same materials but a passive reading task, we replicated the larger LPP in the 300–500 ms time interval for immoral but not moral items and also for negative but not neutral items.

Acknowledgements The work reported in this paper was supported by a grant of the German Research Foundation (DFG) awarded to H.L. (LE 1035/4-1). We thank three anonymous reviewers for their constructive and very helpful comments.

References

- Avramova, Y. R., & Inbar, Y. (2013). Emotion and moral judgment. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4, 169–178.
- Baetens, K., Van der Cruyssen, L., Achtziger, A., Vandekerckhove, M., & Van Overwalle, F. (2011). N400 and LPP in spontaneous trait inferences. *Brain Research*, 1418, 83–92.
- Baggio, G., Van Lambalgen, M., & Hagoort, P. (2008). Computing and recomputing discourse models: An ERP study. *Journal of Memory and Language*, 59, 36–53.
- Bargh, J. A., Chaiken, S., Raymond, P., & Hymes, C. (1996). The automatic evaluation effect: Unconditional automatic attitude activation with a pronunciation task. *Journal of Experimental Social Psychology*, 32, 104–128.
- Bartholow, B. D., Fabiani, M., Gratton, G., & Bettencourt, B. A. (2001). A psychophysiological examination of cognitive processing of and affective responses to social expectancy violations. *Psychological Science*, 12, 197–204.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4 (R Package Version 1.1–13) [Computer software]. Retrieved from <http://CRAN.R-project.org/package=lme4>
- Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7, 1129–1159.
- Bohan, J., Leuthold, H., Hijikata, Y., & Sanford, A. J. (2012). The processing of good-fit semantic anomalies: An ERP investigation. *Neuropsychologia*, 50, 3174–3184.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433–436.
- Brybaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, 58, 412–424.
- Citron, F. M. (2012). Neural correlates of written emotion word processing: A review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain and Language*, 122, 211–226.
- Cunningham, W. A., & Zelazo, P. D. (2007). Attitudes and evaluations: A social cognitive neuroscience perspective. *Trends in Cognitive Sciences*, 11, 97–104.
- Delaney-Busch, N., & Kuperberg, G. (2013). Friendly drug-dealers and terrifying puppies: Affective primacy can attenuate the N400 effect in emotional discourse contexts. *Cognitive, Affective, & Behavioral Neuroscience*, 13, 473–490.
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134, 9–21.
- Donchin, E., & Coles, M. G. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, 11, 357–374.
- Dudschig, C., Mackenzie, I. G., Strozyk, J., Kaup, B., & Leuthold, H. (2016). The sounds of sentences: Differentiating the influence of physical sound, sound imagery, and linguistically implied sounds on physical sound processing. *Cognitive, Affective, & Behavioral Neuroscience*, 16, 940–961.
- Eder, A. B., Leuthold, H., Rothermund, K., & Schweinberger, S. R. (2012). Automatic response activation in sequential affective priming: An ERP study. *Social, Cognitive, and Affective Neuroscience*, 7, 436–445.
- Ferretti, T. R., Singer, M., & Patterson, C. (2008). Electrophysiological evidence for the time-course of verifying text ideas. *Cognition*, 108, 881–888.
- Fields, E. C., & Kuperberg, G. R. (2012). It's all about you: An ERP study of emotion and self-relevance in discourse. *NeuroImage*, 62, 562–574.
- Fields, E. C., & Kuperberg, G. R. (2016). Dynamic effects of self-relevance and task on the neural processing of emotional words in context. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.02003>
- Filik, R., & Leuthold, H. (2008). Processing local pragmatic anomalies in fictional contexts: Evidence from the N400. *Psychophysiology*, 45, 554–558.
- Filik, R., & Leuthold, H. (2013). The role of character-based knowledge in online narrative comprehension: Evidence from eye movements and ERPs. *Brain Research*, 1506, 94–104.
- Fischler, I., Bloom, P. A., Childers, D. G., Arroyo, A. A., & Perry, N. W. (1984). Brain potentials during sentence verification: Late negativity and long-term memory strength. *Neuropsychologia*, 22, 559–568.
- Fischler, I., & Bradley, M. (2006). Event-related potential studies of language and emotion: Words, phrases, and task effects. *Progress in Brain Research*, 156, 185–203.
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293, 2105–2108.
- Günther, F., Dudschig, C., & Kaup, B. (2015). LSafun—An R package for computations based on Latent Semantic Analysis. *Behavior Research Methods*, 47, 930–944.
- Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304, 438–441.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108, 814.
- Hajcak, G., MacNamara, A., & Olvet, D. M. (2010). Event-related potentials, emotion, and emotion regulation: An integrative review. *Developmental Neuropsychology*, 35, 129–155.
- Hajcak, G., Weinberg, A., MacNamara, A., & Foti, D. (2012). ERPs and the study of emotion. In S. J. Luck & E. S. Kappenman (Eds.), *Oxford handbook of ERP components* (pp. 441–474). New York: Oxford University Press.
- Herring, D. R., Taylor, J. H., White, K. R., & Crites, S. L., Jr. (2011). Electrophysiological responses to evaluative priming: The LPP is sensitive to incongruity. *Emotion*, 11, 794–806.
- Holt, D. J., Lynn, S. K., & Kuperberg, G. R. (2009). Neurophysiological correlates of comprehending emotional meaning in context. *Journal of Cognitive Neuroscience*, 21, 2245–2262.
- Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology*, 75, 887–900.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434–446.
- Johnson, R., Jr. (1988). The amplitude of the P300 component of the event-related potential: Review and synthesis. In P. K. Ackles, J. R. Jennings, & M. G. H. Coles, (Eds.), *Advances in psychophysiology* (Vol. 3, pp. 69–138). Greenwich, CT: JAI Press.
- Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3? *Perception*, 36 [ECVP Abstract Supplement, p. 14].
- Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, 1146, 23–49.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621–647.

- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*, 161–163.
- Lai, V. T., Hagoort, P., & Casasanto, D. (2012). Affective primacy vs. cognitive primacy: Dissolving the debate. *Frontiers in Psychology*, *3*, 243. <https://doi.org/10.3389/fpsyg.2012.00243>
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*, 259–284.
- León, I., Díaz, J. M., de Vega, M., & Hernández, J. A. (2010). Discourse-based emotional consistency modulates early and middle components of event-related potentials. *Emotion*, *10*, 863–873.
- Leuthold, H., Filik, R., Mackenzie, I. G., & Murphy, K. (2012). The on-line processing of socio-emotional information in prototypical scenarios: Inferences from brain potentials. *Social, Cognitive, and Affective Neuroscience*, *7*, 457–466.
- Leuthold, H., Kunkel, A., Mackenzie, I. G., & Filik, R. (2015). Online processing of moral transgressions: ERP evidence for spontaneous evaluation. *Social, Cognitive, and Affective Neuroscience*, *10*, 1021–1029.
- Moll, J., de Oliveira-Souza, R., Bramati, I. E., & Grafman, J. (2002). Functional networks in emotional moral and nonmoral social judgments. *NeuroImage*, *16*, 696–703.
- Morris, J. P., Squires, N. K., Taber, C. S., & Lodge, M. (2003). Activation of political attitudes: A psychophysiological examination of the hot cognition hypothesis. *Political Psychology*, *24*, 727–745.
- Nieuwenhuis, S., Aston-Jones, G., & Cohen, J.D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin*, *131*, 510–532.
- Nieuwland, M. S., & Van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, *18*, 1098–1111.
- Nolan, H., Whelan, R., & Reilly, R. (2010). FASTER: Fully automated statistical thresholding for EEG artifact rejection. *Journal of Neuroscience Methods*, *192*, 152–162.
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*(156869). <https://doi.org/10.1155/2011/156869>
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
- Sevinc, G., & Spreng, R. N. (2014). Contextual and perceptual brain processes underlying moral cognition: A quantitative meta-analysis of moral reasoning and moral emotions. *PLOS ONE*, *9*, e87427.
- Storbeck, J., & Clore, G. L. (2007). On the interdependence of cognition and emotion. *Cognition and Emotion*, *21*, 1212–1237.
- Van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, *11*, 657–671.
- Van Berkum, J. J. A., Holleman, B., Nieuwland, M., Otten, M., & Murre, J. (2009). Right or wrong? The brain's fast response to morally objectionable statements. *Psychological Science*, *20*, 1092–1099.
- Van de Meerendonk, N., Kolk, H. H. J., Chwilla, D. J., & Vissers, C. Th. W. M. (2009). Monitoring in language perception. *Language and Linguistics Compass*, *3*, 1211–1224.
- Van Duynslaeger, M., Van Overwalle, F., & Verstraeten, E. (2007). Electrophysiological time course and brain areas of spontaneous and intentional trait inferences. *Social, Cognitive, and Affective Neuroscience*, *2*, 174–188.
- Xiang, M., & Kuperberg, G. (2015). Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience*, *30*, 648–672.
- Zhang, Q., Lawson, A., Guo, C., & Jiang, Y. (2006). Electrophysiological correlates of visual affective priming. *Brain Research Bulletin*, *71*, 316–323.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*, 162–185.