doi:10.1093/scan/nsu151 SCAN (2015) 10, 1021–1029

Online processing of moral transgressions: ERP evidence for spontaneous evaluation

Hartmut Leuthold,[1] Angelika Kunkel,[1] Ian G. Mackenzie,[1] and Ruth Filik[2]
[1]Department of Psychology, University of Tübingen, Schleichstr. 4, 72076 Tübingen, Germany and [2]School of Psychology, University of Nottingham, University Park, Nottingham NG7 2RD, UK

Experimental studies using fictional moral dilemmas indicate that both automatic emotional processes and controlled cognitive processes contribute to moral judgments. However, not much is known about how people process socio-normative violations that are more common to their everyday life nor the time-course of these processes. Thus, we recorded participants' electrical brain activity while they were reading vignettes that either contained morally acceptable *vs* unacceptable information or text materials that contained information which was either consistent or inconsistent with their general world knowledge. A first event-related brain potential (ERP) positivity peaking at ~200 ms after critical word onset (P200) was larger when this word involved a socio-normative or knowledge-based violation. Subsequently, knowledge-inconsistent words triggered a larger centroparietal ERP negativity at ~320 ms (N400), indicating an influence on meaning construction. In contrast, a larger ERP positivity (larger late positivity), which also started at ~320 ms after critical word onset, was elicited by morally unacceptable compared with acceptable words. We take this ERP positivity to reflect an implicit evaluative (good–bad) categorization process that is engaged during the online processing of moral transgressions.

Keywords: moral transgression; socio-normative evaluation; P200; N400; LPP

INTRODUCTION

Imagine you are told that you can save the life of five people working on a track from being run over by a trolley by pushing a man from a footbridge onto the track. Whereas this stops the trolley, of course, you cause the death of an innocent person. Would you do it? This vignette illustrates a common procedure used to examine the nature of human moral judgment by presenting a short story, in this case a moral dilemma, to reveal the processes and principles underlying such judgments (for a review, see Waldmann *et al.*, 2012). Whereas it is certainly illuminating to study the behavioral and brain responses triggered by such extreme fictional situations, in order to understand our everyday moral and, more generally, socio-normative evaluations, it is important to complement this approach by one that targets situations that we encounter more commonly in our social environment (e.g. cheating, telling lies, etc.; cf. Haidt and Kesebir, 2010). Such everyday situations involving moral transgressions are likely to be rapidly evaluated on the basis of matching personal experiences and social knowledge stored in episodic and semantic memory (e.g. Tulving, 2002; Blank, 2009). Therefore, in this study, we will investigate these evaluative processes with millisecond temporal resolution by using event-related brain potentials (ERPs).

Recent accounts of moral judgment have rejected the long-held cognitive-developmental assumption that moral judgments reflect the output of a rational reasoning system (Kohlberg, 1981). Thus, according to Haidt (2001) social intuitivist model (SIM), people decide and act based on moral intuitions, which he defines as 'fast, automatic and (usually) affect-laden processes in which an evaluative feeling of good–bad or like–dislike (about the actions or character of a person) appears in consciousness [. . .]' (Haidt, 2007, p. 998). According to SIM, a person's moral intuitions have primacy and provide the grounds for deliberate moral reasoning, such that controlled reasoning processes typically support the construction of post-hoc justifications for intuition-based judgments and actions.

Based on functional magnetic resonance imaging (fMRI) evidence, Greene *et al.* (2001, 2004) proposed that both automatic emotional processes and controlled cognitive processes determine moral judgments. More specifically, they assume that emotional processes trigger deontological judgments (adhering to moral duties), whereas slower controlled processes trigger utilitarian judgments (maximizing expected utility). Greene *et al.* (2001) found that when so-called personal moral dilemmas were presented that involved physical contact with the person who is harmed, as in the above footbridge version of the 'trolley dilemma' (where you would have to physically push the person onto the tracks), brain regions concerned with emotional processing and social cognition (posterior cingulate gyrus, medial prefrontal cortex) were more strongly activated. In contrast, impersonal dilemmas, like the classic trolley dilemma where five people can be saved by flipping a switch that redirects the out-of-control trolley onto a side track where it runs over a single person instead, elicited stronger neural activity in brain regions concerned with working memory and deliberate reasoning (dorsolateral prefrontal cortex, inferior parietal lobe). Also in accord with dual-process contributions was the finding that for personal dilemmas, judgment times were longer when participants chose the utilitarian (kill one person to save more lives) rather than the deontological option (killing is wrong). This finding suggested that the automatic emotional process primes a deontological response, whereas the controlled process activates the utilitarian response, thereby producing a conflict that takes time to resolve (also see behavioral studies of Greene *et al.*, 2008; Suter and Hertwig, 2011; Conway and Gawronski, 2013).

Together, behavioral and fMRI studies have provided crucial insights into the processes underlying moral judgment (but see Kahane, 2012). Critically, due to the use of explicit moral judgment tasks and the limited temporal resolution of fMRI, it remains largely unclear (i) how rapidly moral evaluations take place and (ii) whether the associated processes are triggered without the participants being asked to explicitly evaluate the materials. We believe that ERPs, with their millisecond temporal resolution and ability to monitor linguistic processes during reading without the need for a secondary task, are ideally suited

Received 9 January 2014; Accepted 24 December 2014
Advance Access publication 1 January 2015
HL was supported by funding from the German Scholars Organization (GSO, Berlin) through a '"German Researcher Relocation'" program grant of the Alfried Krupp von Bohlen und Halbach Foundation.
Correspondence should be addressed to Hartmut Leuthold, Department of Psychology, Eberhard Karls University of Tübingen, Schleichstr. 4, 72076 Tübingen, Germany. E-mail: hartmut.leuthold@uni-tuebingen.de

to address both issues. In addition, previous studies suggest that specific ERP components can be informative regarding the evaluative processing of moral information, as we will briefly review below.

ERPS IN MORAL DECISION-MAKING

Given the clear importance of the time-course of processing, it is somewhat surprising that only three ERP studies, at least to our knowledge, have investigated moral decision-making. Chen *et al.* (2009) tested Chinese participants who had experienced the Sichuan earthquake in May 2008. They presented name-pairs that referred either to relatives (e.g. father–mother) or to strangers (e.g. stranger A—stranger B) and asked participants to decide which of the two people they would rescue in the case of an earthquake. In this moral dilemma task, they observed a larger frontocentral positivity (P200), peaking at ~190–200 ms, followed by a larger centroparietal late positivity [P300 or larger late positivity (LPP); 350–450 ms] when the choice was between two relatives rather than between two strangers, although the LPP effect occurred only after participants were (falsely) informed about a forthcoming aftershock halfway through the experiment. The authors proposed that the P200 reflects the initial evaluation of the dilemma and conflict detection, whereas the LPP relates to dilemma resolution during moral decision-making.

However, an interpretation unrelated to moral decision-making appears conceivable given that the two conditions in Chen *et al.*'s study differed with respect to the familiarity of the names and hence, among other things, their emotional salience and the attitudes participants presumably held toward them. Specifically, it is known that emotional compared with neutral word stimuli trigger a larger P200 component and a larger LPP (cf. Kissler *et al.*, 2006). Also, explicit and implicit evaluative processing of contextually incongruent attitudinal stimuli triggers a larger centroparietal LPP (e.g. Cacioppo *et al.*, 1993; Crites *et al.*, 1995; Ito and Cacioppo, 2000). Thus, the P200 and LPP effects observed by Chen *et al.* might relate, respectively, to the attention-capturing nature of emotional (arousing) stimuli and the subsequent evaluative categorization of incoming information.

Sarlo *et al.* (2012) employed moral dilemmas that differed with regard to whether the death of one or more persons was intended to save many others (instrumental dilemma—e.g. footbridge dilemma) *vs* a foreseen but unintended side effect (incidental dilemma—e.g. trolley dilemma). Upon presentation of a decision screen, participants were to choose between a utilitarian (one person is intentionally killed to save many) *vs* non-utilitarian option (many die but no-one is intentionally killed). ERPs time-locked to the decision screen showed a larger positivity over anterior electrodes (P260) for instrumental than incidental dilemmas, whereas a later posterior positivity (600–750 ms) was larger following incidental than instrumental dilemmas. In line with Greene *et al.*'s (2001) dual-process theory, Sarlo *et al.* suggested that instrumental moral dilemmas trigger a rapid affective evaluative response (P260), whereas incidental dilemmas demand more controlled processing (late positivity). It should be noted though that this study was not concerned with the evaluative processes that are triggered during the reading of the moral dilemmas, but instead by the later decision.

Recently, Van Berkum *et al.* (2009) addressed this issue by examining whether and when the values held by an individual influence the linguistic analysis of meaning. Two groups of male participants with opposing value systems (non-Christians *vs* Christians) were asked to read statements that contained a critical evaluative word (e.g. 'In a bad marriage, divorce is an *acceptable/unacceptable* solution'; italics added here) and subsequently to rate their agreement. In addition to P200 and LPP time intervals, they analyzed the language-based N400 component, which manifests as a centroparietally distributed, negative-going deflection in the ERP with an onset around 200 ms and a peak at ~400 ms (for a review, see Kutas and Federmeier, 2011). Based on the finding that larger N400s are elicited by words that are unpredictable, a poor fit with context, or violate the participant's world knowledge than those that are a good fit, are more predictable, or accord with world knowledge (e.g. Kutas and Hillyard, 1984; Hagoort *et al.*, 2004; DeLong *et al.*, 2005; Van Berkum *et al.*, 1999, 2005; Sanford *et al.*, 2011; Filik and Leuthold, 2008, 2013), it is assumed that the N400 reflects the ease of constructing the meaning of an incoming word within the sentence or discourse context given (cf. Kutas and Federmeier, 2011; Van Berkum, 2012). Van Berkum *et al.* found that value-inconsistent words (e.g. acceptable for strict-Christians *vs* unacceptable for non-Christians) initially elicited a larger, widely distributed positivity between 200 and 250 ms (P200), followed by a larger centroparietal negativity (N400) between 375 and 425 ms and finally a LPP between 500 and 650 ms. According to Van Berkum *et al.*, the N400 findings indicate that readers immediately and automatically evaluate incoming information with respect to their personally held values, giving rise to rapid, valence-based influences on meaning construction. They further speculate that this N400 effect cancels a single sustained LPP effect with an earlier onset than the N400, therefore emerging as a larger P200 and LPP for value-inconsistent than value-consistent statements, an ERP signature they take to reflect the activation of the affect system.

Together, the three ERP studies investigating the neural processes associated with moral judgments have shown larger ERP positivities starting at ~200 ms, potentially relating to the immediate affective evaluation of incoming information. In addition, the N400 findings of Van Berkum *et al.* accord with other studies suggesting an influence of socio-emotional factors on meaning construction. For example, it has been found that stereotype violations (Van Berkum *et al.*, 2008; White *et al.*, 2009) as well as socio-emotional expectancy violations (León et *al.*, 2010; Leuthold *et al.*, 2012) compared with non-violations trigger an N400 effect.

It remains unclear, however, whether the reviewed ERP findings translate to moral transgressions (or socio-normative scenarios) that are more commonly encountered in everyday life, in particular when no explicit (moral or agreement) judgments are required [as Van Berkum *et al.* (2009) also pointed out]. In addition, regarding the P200 effect, its functional interpretation as reflecting an initial affective evaluation awaits clarification. The reason for this is that larger P200s have been observed under various conditions that point to a different interpretation: (i) during visual feature detection and attentional processing of rare target stimuli (e.g. Hillyard and Münte, 1984; Luck and Hillyard, 1994), (ii) in the social cognition domain when expectations about a character's behavioural goals, as inferred from the story context, are violated (e.g. Van der Cruyssen *et al.*, 2009) and to racial outgroup stimuli (e.g. Dickter and Bartholow, 2007) and (iii) in the language domain when good-fit semantic anomalies are detected rather than missed (Bohan *et al.*, 2012). Together, these findings appear to suggest that valenced, rare or unexpected stimuli capture attention early in the time-course of visual information processing. Hence, it seems likely that the P200 component is associated with the attentional processing of (visual) input rather than its affective evaluation.

In contrast, the LPP component seems to reflect a more specific, task-independent (implicit) evaluative categorization process. This conjecture accords with the earlier mentioned findings of Cacioppo and colleagues (Cacioppo *et al.*, 1993; Crites *et al.*, 1995), who showed that LPP amplitude is enhanced by evaluative inconsistencies among stimuli, even when the task does not demand their explicit processing (e.g. Ito and Cacioppo, 2000). It is also consistent with the observation that evaluatively incongruent prime-target pairs (either pictures or

Table 1 Example socio-normative and general world knowledge materials (critical word in italics)

Socio-normative	
Morally acceptable context	Tinas Opa leidet an Krebs und wird bald sterben. Zu seinem 85. Geburtstag hat er ein großes Fest geplant und wünscht sich vor allem, dass alle seine Kinder und Enkel kommen. [Tina's grandfather suffers from cancer and will die soon. For his 85th Birthday he has planned a big party and wishes nothing more than that all his children and grandchildren attend.]
Morally unacceptable context	Tinas Chef macht schon seit einer Weile eindeutige Anspielungen. Nun hat er sie zu einem Essen in ein teures Restaurant eingeladen. Sie weiß, dass er seit 20 Jahren verheiratet ist und drei Kinder hat. [Tina's boss has been making explicit innuendos for some time. Now he has invited her out for dinner to an expensive restaurant. She knows that he has been married for 20 years and is the father of three children.]
Target sentence[a]	Sie hat die Einladung *angenommen*. [She has the invitation *accepted*.]
World knowledge	
Knowledge-consistent context	Bei einem Frankreichaustausch isst Frau Lehmann eine bekannte französische Delikatesse. [During a France exchange Mrs. Lehmann eats a famous French speciality.]
Knowledge-inconsistent context	Frau Lehmann geht in ein schwäbisches Restaurant und bestellt eine lokale Spezialität. [Mrs. Lehmann goes to a Schwabian Restaurant and orders a local speciality.]
Target sentence[a]	Sie erhält als Gericht einen Teller voller *Schnecken* und Weißbrot. [She receives as dish a plate full of *snails* and white bread.]

[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.

words) trigger a larger LPP but no N400 effect (Herring *et al.*, 2011). The finding that individual attitude differences (e.g. racial bias) were reflected in LPP amplitude but not in self-reports (e.g. Ito *et al.*, 2004) further indicates the LPP's sensitivity to implicit evaluative processes. Moreover, social cognition studies using language materials to examine intentional and incidental traits or goal inferences observed larger LPP (P300) amplitudes to critical words that indicated trait- or goal-inconsistent behavior (e.g. Bartholow *et al.*, 2001; Van Duynslaeger *et al.*, 2007; Van der Cruyssen *et al.*, 2009).

In summary, the recording and analysis of ERPs seems particularly suited to produce novel insights regarding the time-course and mechanisms underlying implicit socio-normative judgments. That is, it is evident that ERPs give access to covert processes at different levels, which manifest themselves in distinct ERP components reflecting the attentive processing of incoming information (P200), the construction of meaning (N400) and the implicit evaluation of input (LPP).

OBJECTIVES OF THE PRESENT STUDY

In this article, participants were tested in a text comprehension task which complements previous research on moral judgment (i) by demanding no explicit evaluative judgments, hence, allowing us to study implicit evaluative processes and (ii) by employing vignettes that describe subjectively more probable situations (cf. Haidt and Kesebir, 2010).

We created prototypical scenarios (see Table 1) in which an initial sentence was used to establish the social context. The test sentence then contained a critical word that described a behavior that was either morally acceptable or morally unacceptable relative to the context. Critically, to eliminate possible word-related effects due to the use of target words differing in valence, only the wording of the context but not that of the critical test sentences varied between conditions.

In order to check for the specificity of ERP effects associated with socio-normative evaluations, we included additional materials that differed with respect to a morality-unrelated text dimension. For these additional materials, contexts were created which were followed by a target sentence containing a critical word that was either consistent or inconsistent with the participants' general world knowledge, or simply consistent or inconsistent with the context in which they appeared (see Table 1). For the sake of brevity, henceforth, we refer to these materials as 'knowledge-based materials' and to the experimental conditions as 'knowledge-consistent' or 'knowledge-inconsistent' conditions.

Based on the earlier reviewed ERP findings, we predict that morally unacceptable compared with acceptable items will trigger a larger early ERP positivity (P200; Chen *et al.*, 2009; Van Berkum *et al.*, 2009), whereas world knowledge violations will not influence the P200 (e.g. Filik and Leuthold, 2013). This finding would support the assumption of a special role of socio-normative information in capturing attention, thereby undergoing enhanced processing. In addition, based on previously reported N400 effects, we anticipate that larger N400 amplitudes will be triggered by both morally unacceptable than acceptable words (Van Berkum *et al.*, 2009; Leuthold *et al.*, 2012) and knowledge-inconsistent than consistent words (e.g. Hagoort *et al.*, 2004; Filik and Leuthold, 2008, 2013). Finally, we further expect the LPP to reflect the implicit evaluative categorization (e.g. Ito and Cacioppo, 2000; Van der Cruyssen *et al.*, 2009) of socio-normative materials but not of knowledge-based materials (e.g. Van de Meerendonk *et al.*, 2010).

METHOD

Participants

Thirty-four right-handed native German speakers from the University of Tübingen [24 females, $M = 25.02$ years, standard error of mean (SEM) = 0.55 years] received course credits or payment for participating. Data from six participants were excluded from the analyzes due to excessive alpha activity ($N = 3$), substantial drifts in electroencephalogram (EEG) activity at multiple electrodes ($N = 2$) or due to <50% of trials per condition remaining after artifact correction and rejection ($N = 1$).

Materials and design

Experimental materials consisted of 80 items for the socio-normative conditions and a further 80 items for the knowledge-based conditions (see Table 1 for examples; the full set is available from the first author). Each item consisted of two parts. The first part was a context consisting of two or three sentences, and the second part was the target sentence that contained the critical word. Socio-normative items described behaviors that would either be morally acceptable or unacceptable, whereas knowledge-based materials would contain a general world knowledge-consistent or a knowledge-inconsistent word. In addition, 40 filler items were constructed which contained no inconsistencies and were of a similar length to the experimental items.

Randomization of items and conditions across participants was performed in the following way. For every two participants, two lists of items were randomly generated anew, one for each participant.

In a given list, each socio-normative item appeared either in the morally acceptable or the morally unacceptable condition, and each knowledge-based item either in the knowledge-consistent or the knowledge-inconsistent condition but appeared in both socio-normative and both knowledge conditions across the two lists. Each 200-item list consisted of 40 morally acceptable and 40 morally unacceptable items, 40 knowledge-consistent and 40 knowledge-inconsistent items, as well as 40 filler items. The order of items within each list was randomized with the following constraints: the maximum number of same condition repetitions was three items, with the transitional (sequence) probability between conditions being approximately equal.

Pre-tests of materials

Full information about the pre-testing of materials and complete results are provided in the Supplementary Data. Statistical analysis of rating data showed that morally acceptable items were rated as being more morally acceptable than morally unacceptable items ($M = 5.89$ *vs* 2.70), $t(79) = 21.04$, $P < 0.001$, and had slightly higher plausibility scores ($M = 6.00$ *vs* 5.33), $t(79) = 4.38$, $P < 0.001$.

Procedure

Following electrode application, participants sat facing a computer screen with their chin resting on a chin rest, in order to minimize any head movements. Experimental materials were presented using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner *et al.*, 2007) running under MATLAB 2012a, on a Mac Mini OS 10.7. Participants were instructed to read the stories, and to answer comprehension questions at certain times during the experiment by pressing the appropriate response button. They were asked to maintain fixation at the center of the screen during word presentation.

During each trial, participants started presentation of the complete context sentences by pressing the 'Proceed' button. The context sentences were displayed for a minimum duration of 1500 ms. When participants had read the context sentences they pressed again the 'Proceed' button to initiate word-by-word presentation of the target sentence. Following the presentation of a fixation point for 1000 ms, each word was displayed centrally for 300 ms, with a 200-ms blank interval between successive word presentations. After the final word was presented, there was a blank interval of 2000 ms. One in 10 of the items of each block was followed by a verification statement that required a 'true' or 'false' response via a button press, in order to ensure that participants were attending to the materials. For items followed by the verification statement, as soon as their response was given, the screen instructing participants to press the space bar for presentation of the next discourse appeared. The mean correct response rate for verification statements was 78.9% (SEM = 1.68), indicating that they were reading for comprehension.

The test items were presented in a total of 20 blocks. Each block consisted of 10 items, and blocks were separated by a short break that was controlled in its duration by the participant.

Electrophysiological measures

EEG activity was recorded continuously from 72 Ag-AgCl electrodes using a BioSemi Active-Two amplifier system with a sampling rate of 512 Hz. EEG data were checked for artifacts and corrected (see Supplementary Data for details) following a procedure similar to that described by Nolan *et al.* (2010). There remained on average 36.64 trials (out of 40; range = 18–40, median = 38) per condition.[1]

[1] For all participants and conditions, there were ≥ 28 trials included in each condition specific average ERP waveform except for a single participant including fewer trials in congruent ($N = 21$) and incongruent ($N = 18$) world-knowledge conditions. ANOVAs with this participant excluded replicated the reported results.

Data analysis

All EEG/ERP analysis was performed using available MATLAB toolboxes (EEGLAB: Delorme and Makeig, 2004; FieldTrip: Oostenveld *et al.*, 2011) and custom MATLAB scripts. The analysis epoch started 500 ms prior to the onset of the critical word and lasted until 1500 ms after it. 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 recalculated to an average mastoid reference. Averaged ERPs were low-pass filtered (25 Hz, 36 dB/oct) and aligned to a 200-ms baseline prior to the onset of the critical word. Mean ERP amplitudes were determined for P200 (200–250 ms), N400 (300–500 ms) and LPP (500–1000 ms) time ranges. In addition, we conducted separate non-parametric cluster-based permutation analyzes to determine the onset of moral acceptability and knowledge-based effects in the ERP waveform (cf. Supplementary Data).

Statistical analyzes were performed by means of Huynh–Feldt corrected repeated measures analyzes of variance (ANOVA), separately for ERP amplitudes at midline electrodes (AFz to Oz) and at lateral electrode sites (defined by eight regions of interest—ROI; see Supplementary Data for details). For the analysis of ERP amplitude data recorded from midline electrodes, we performed an ANOVA with materials (socio-normative *vs* knowledge-based), condition (morally acceptable *vs* unacceptable or knowledge-consistent *vs* inconsistent) and electrode (AFz, Fz, FCz, Cz, CPz, Pz, Poz, Oz) as variables. For the analysis of ERP amplitudes recorded from lateral electrode sites, we performed an ANOVA with materials, condition, hemisphere (left, right), ant-pos (anterior, posterior) and verticality (ventral, dorsal) as variables. In the case of significant Material x Condition (x Electrode/ROI) interactions, separate statistical analyzes were carried out for knowledge-based and socio-normative materials. For all statistical analyzes, the significance level was set to alpha = 0.05.

RESULTS

Figures 1 and 2 display the grand average ERP waveforms triggered by knowledge-consistent *vs* inconsistent words and by morally acceptable *vs* unacceptable words, respectively. The topographic distributions of the knowledge violation and the moral acceptability effect are shown in Figure 3 (top *vs* bottom panel). As can be seen in Figure 2, starting at ~200 ms after word onset, the ERP revealed a generally more positive-going waveform for morally unacceptable than acceptable sentences, which is in clear contrast with the N400 effect observed for knowledge-based materials (cf. Figures 1 and 3).

200–250 ms (P2)

In this time window, averaged midline ERP amplitudes were more positive for knowledge-based than for socio-normative materials (5.51 *vs* 4.78 μV), $F(1, 27) = 6.35$, $P < 0.05$, $\eta_p^2 = 0.19$, and this material effect was maximal over centroparietal electrodes as indicated by the Material x Electrode interaction, $F(7, 189) = 6.69$, $P < 0.001$, $\varepsilon = 0.45$, $\eta_p^2 = 0.20$. The significant main effect of condition, $F(1, 27) = 6.40$, $P < 0.05$, $\eta_p^2 = 0.19$, indicated a larger ERP positivity for knowledge-inconsistent/morally unacceptable than consistent/acceptable items (5.43 *vs* 4.87 μV). Although midline ERP amplitude appeared to be numerically more positive-going for morally unacceptable than morally acceptable items (5.27 *vs* 4.30 μV), but hardly so for knowledge-inconsistent than consistent items (5.59 *vs* 5.44 μV), the Material x Condition interaction was not significant, $F(1, 27) = 2.32$, $P = 0.14$, $\eta_p^2 = 0.08$, and also not in interaction with electrode, $F(7, 189) = 0.20$, $P = 0.29$. The analysis of ERP amplitudes at lateral ROIs confirmed the main effect of material, $F(1, 27) = 5.73$, $P < 0.05$, $\eta_p^2 = 0.18$, and this effect was again stronger over posterior ROIs, $F(1, 27) = 20.39$, $P < 0.001$, $\eta_p^2 = 0.43$ (cf. Figure 3).

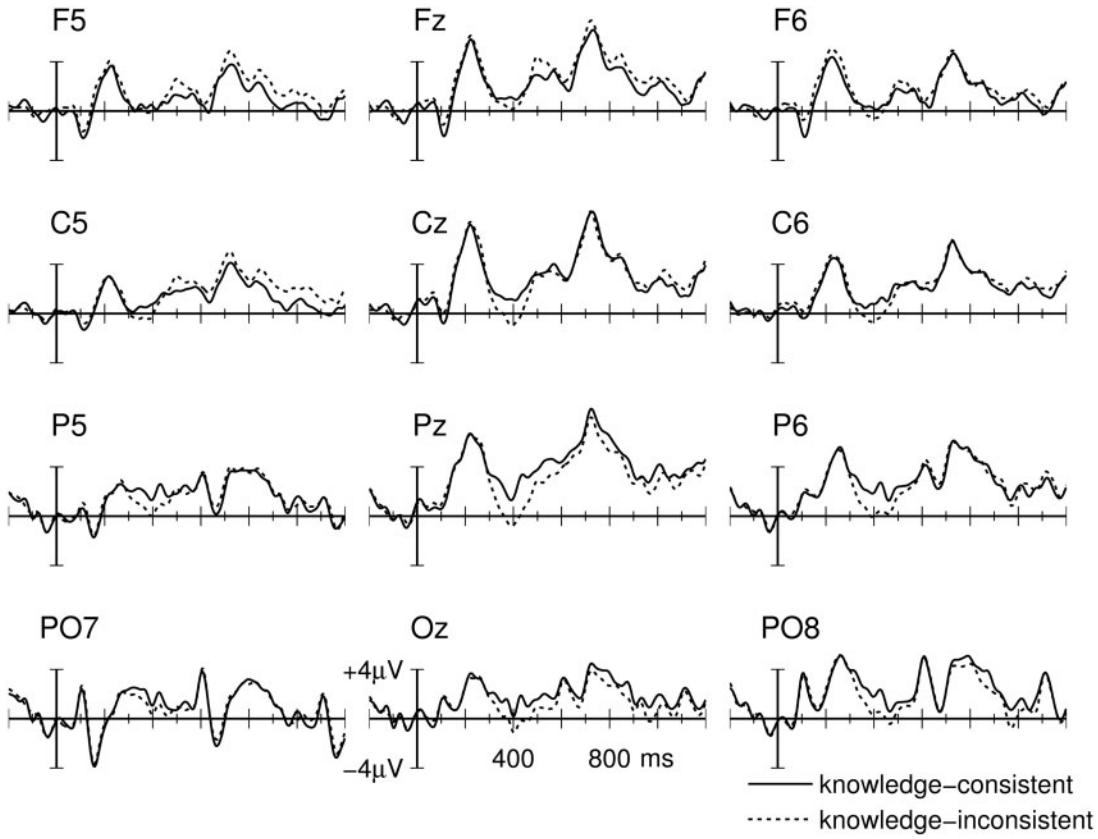

Fig. 1 Grand average ERP waveforms elicited at midline and lateral electrodes time-locked to the onset of the critical word for world knowledge materials. Positivity is plotted upwards.

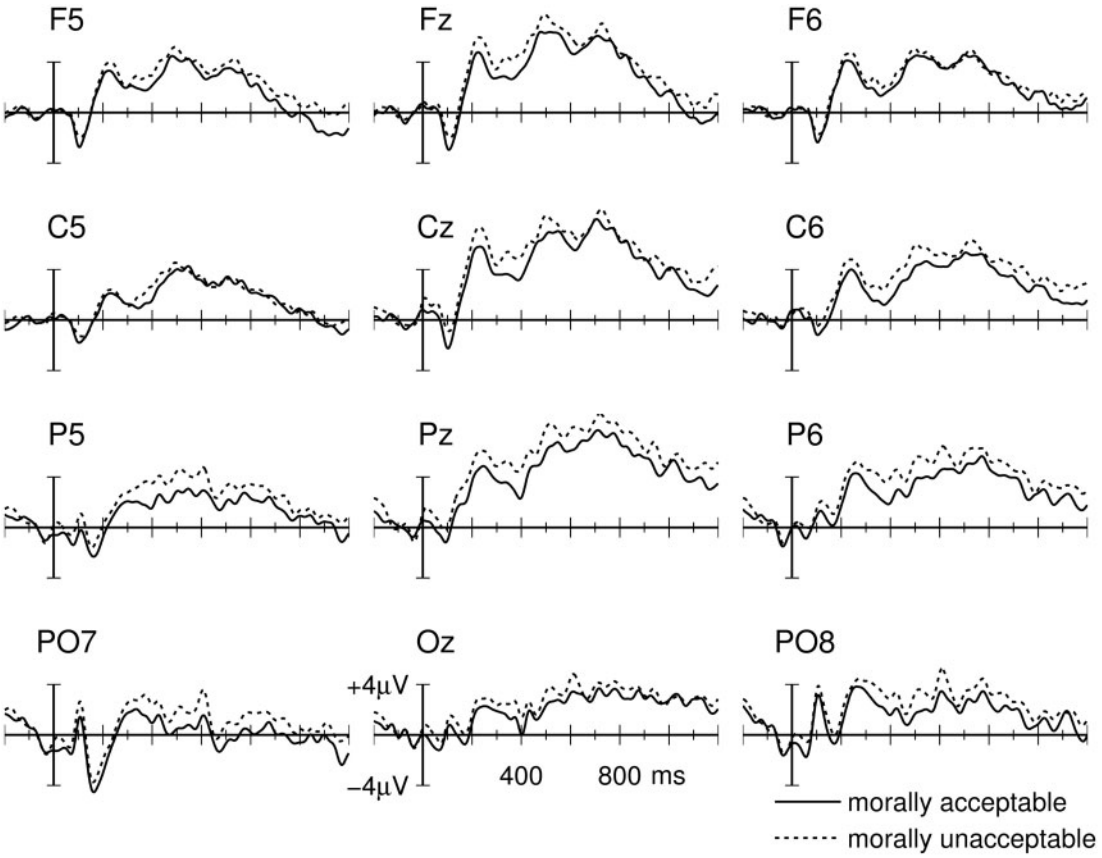

Fig. 2 Grand average ERP waveforms elicited at midline and lateral electrodes time-locked to the onset of the critical word for socio-normative materials. Positivity is plotted upwards.

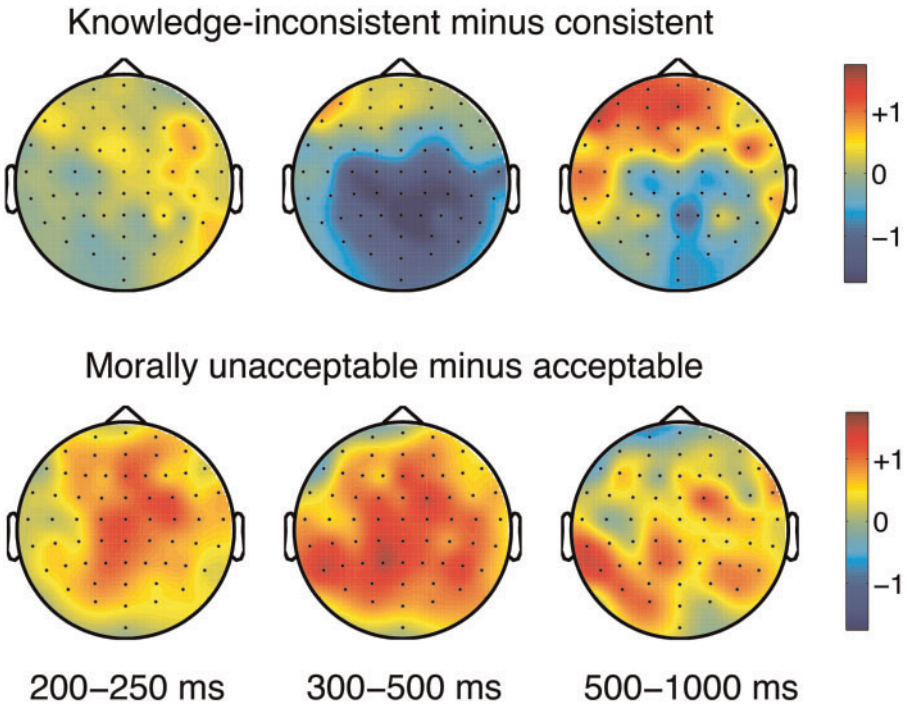

Fig. 3 Spline-interpolated topographic map of mean ERP difference waveform for the 200–250, 300–500 and 500–1000-ms time interval. Top panel: world knowledge condition (knowledge-inconsistent minus knowledge-consistent). Bottom panel: socio-normative condition (morally unacceptable minus morally acceptable).

300–500 ms

In this time window, averaged midline ERP amplitudes were more positive for socio-normative than for knowledge-based materials (4.08 *vs* 1.15 μV), $F(1, 27) = 92.17$, $P < 0.001$, $\eta_p^2 = 0.77$. Importantly, there was a reliable Material x Condition interaction, $F(1, 27) = 12.96$, $P < 0.01$, $\eta_p^2 = 0.32$, which was further modulated by electrode, $F(7, 189) = 4.40$, $P < 0.001$, $\varepsilon = 0.36$, $\eta_p^2 = 0.32$. Similarly, the analysis of ERP amplitudes at lateral ROIs revealed a main effect of material, $F(1, 27) = 69.20$, $P < 0.001$, $\eta_p^2 = 0.72$, and a significant Material x Condition interaction that was modulated by topographic factors, all $Fs > 4.26$, $ps < 0.05$.

The separate analysis of knowledge-based materials revealed that in the N400 time interval, knowledge-inconsistent compared with consistent words elicited a more negative-going ERP waveform over midline electrodes (0.66 *vs* 1.65 μV), $F(1, 27) = 4.09$, $P = 0.05$, $\eta_p^2 = 0.13$. This N400 knowledge violation effect was maximal over centroparietal electrodes, $F(7, 189) = 8.20$, $P < 0.001$, $\varepsilon = 0.47$, $\eta_p^2 = 0.23$ (cf. Figure 3). The analysis of ERP amplitudes at lateral ROIs corroborated a stronger knowledge violation effect over posterior than anterior ROIs, $F(1, 27) = 12.51$, $P < 0.01$, $\eta_p^2 = 0.32$ (cf. Figure 3).

The analysis of socio-normative materials showed a midline ERP positivity with a central-maximal distribution, $F(7, 189) = 7.13$, $\varepsilon = 0.31$, $P < 0.05$, $\eta_p^2 = 0.21$. Crucially, this positivity was larger for morally unacceptable than acceptable items (4.62 *vs* 3.53 μV), $F(7, 189) = 22.61$, $P < 0.001$, $\eta_p^2 = 0.46$. The Condition x Electrode interaction was not significant, $F(7, 189) < 1$, $P = 0.86$.

The lateral ROI analysis confirmed the main effect for the morally unacceptable *vs* morally acceptable conditions (3.36 *vs* 2.45 μV), $F(1, 27) = 7.41$, $P < 0.05$, $\eta_p^2 = 0.22$. In addition, the Condition x Hemisphere x Ant-Pos interaction was significant, $F(1, 27) = 6.48$, $P < 0.05$, $\eta_p^2 = 0.19$, due to a reliable moral acceptability effect over right anterior and posterior ROIs (3.67 *vs* 2.78 μV and 3.30 *vs* 2.33 μV, respectively) as well as over left posterior ROIs (2.99 *vs* 1.83 μV), all $Fs(1, 27) \geq 5.00$, $ps < 0.05$, but not over left anterior ROIs (3.49 *vs* 2.84 μV), $F(1, 27) = 2.14$, $P = 0.16$.

500–1000 ms

In this time window, averaged midline ERP amplitudes were more positive for socio-normative than for knowledge-based materials (5.22 *vs* 3.57 μV), $F(1, 27) = 13.03$, $P < 0.01$, $\eta_p^2 = 0.33$. The Material x Electrode interaction was significant, $F(7, 189) = 2.15$, $P < 0.05$, $\varepsilon = 0.50$, $\eta_p^2 = 0.07$, reflecting the stronger material effect over centroparietal electrodes. In addition, the Condition x Electrode interaction, $F(7, 189) = 3.04$, $P < 0.01$, $\varepsilon = 0.51$, $\eta_p^2 = 0.10$, and the Material x Condition x Electrode interaction were significant, $F(7, 189) = 5.03$, $P < 0.001$, $\eta_p^2 = 0.16$. The analysis of ERP amplitudes at lateral ROIs confirmed the main effect for socio-normative *vs* knowledge-based materials, $F(1, 27) = 6.19$, $P < 0.05$, $\eta_p^2 = 0.17$. Also, the Material x Condition x AntPost interaction was significant, $F(1, 27) = 9.63$, $P < 0.01$, $\eta_p^2 = 0.26$.

Material-specific analyzes indicated a significant Condition x Electrode interaction for knowledge-based materials, $F(7, 189) = 7.82$, $\varepsilon = 0.55$, $P < 0.001$, $\eta_p^2 = 0.22$. As in the preceding time interval, this interaction was due to a more negative-going ERP waveform over centroparietal electrodes in the knowledge-inconsistent than the knowledge-consistent condition. The analysis of mean ERP amplitudes at lateral sites produced a reliable Condition x Hemisphere, $F(1, 27) = 6.79$, $P < 0.05$, $\eta_p^2 = 0.20$, and Condition x Ant-Pos interaction, $F(1, 27) = 7.55$, $P < 0.05$, $\eta_p^2 = 0.22$, however, separate ROI-dependent analyzes revealed no significant condition effects, all $Fs < 2.58$, $ps > 0.12$.

The analysis of socio-normative materials showed midline ERP amplitude not to be reliably influenced by moral acceptability, all $Fs(1, 27) < 1.35$, $ps > 0.25$. The analysis of mean ERP amplitudes at lateral ROIs produced a significant Condition x Ant-Pos x Hemisphere interaction, $F(1, 27) = 7.34$, $P < 0.05$, $\eta_p^2 = 0.21$, due to

a reliable moral acceptability effect only over left posterior ROIs, $Fs(1, 27) = 4.59$, $P < 0.05$, but not over right posterior ROIs, $Fs(1, 27) = 3.34$, $P = 0.08$ or anterior ROIs, $Fs(1, 27) < 1.29$, $ps > 0.27$.

ERP onset latency

For knowledge-based materials, the cluster-based permutation test revealed a significantly more negative ERP amplitude for knowledge-inconsistent than consistent items beginning at 322 ms (cluster 322–453 ms, $P < 0.05$). For socio-normative materials, morally unacceptable compared with acceptable items produced a more positive ERP waveform starting at 318 ms (cluster 318–435 ms, $P < 0.05$).

DISCUSSION

In this study, we investigated the time-course of socio-normative information processing by analyzing ERP correlates known to reflect immediate, covert evaluative categorization processes. The major novel finding was that critical words elicited a broadly distributed positivity that was more pronounced for morally unacceptable than for morally acceptable sentences starting already at ~320 ms after word onset. In contrast, for the general world knowledge-based materials, knowledge-inconsistent as compared with consistent sentences elicited a larger N400-like component that began at the same time, whereas ERP positivities were unaffected. In addition, an ERP positivity peaking at ~200 ms (P200) indicated the enhanced processing of knowledge-based compared with socio-normative materials. Together, the present ERP results indicate that readers process incoming information quite differently for text materials that include socio-normative *vs* knowledge-based violations.

A first, material-specific, ERP effect is the presence of a larger positivity (P200) already ~200–250 ms after the onset of the critical word for knowledge-based than socio-normative materials. This may suggest that the former materials are more salient and hence more intensely processed. In line with this interpretation, larger P200 amplitudes have been reported to be triggered by various experimental conditions, as in the case of processing emotional (e.g. Kissler *et al.*, 2006) and racial outgroup stimuli (e.g. Dickter and Bartholow, 2007), of information that violates moral values (e.g. Chen *et al.*, 2009; Van Berkum *et al.*, 2009), as well as of attended or rare target features (Hillyard and Münte, 1984; Luck and Hillyard, 1994). Moreover, P200 amplitude was larger if the critical word violated discourse-based expectations. In contrast to our predictions, and despite the numeric trend in the expected direction, this violation-based P200 effect was statistically not reliably larger for socio-normative than knowledge-based materials. Therefore, we interpret this P200 effect in terms of the attention-capturing nature of unexpected (or rare) linguistic input, be it related to moral transgressions or world knowledge.

Following the P200, for knowledge-based materials a larger negative-going deflection was observed between 300 and 500 ms after the onset of knowledge-inconsistent than consistent words, and this effect continued during the 500–1000-ms time interval. Given the topographic distribution and time-course of this ERP effect, we assume that it reflects a knowledge (or plausibility) violation-based N400 effect, as in similar previous studies (e.g. Hagoort *et al.*, 2004; Van de Meerendonk *et al.*, 2010; Sanford *et al.*, 2011; Bohan *et al.*, 2012; Filik and Leuthold, 2008, 2013). This particular result suggests that knowledge-based expectations are indeed triggered by the context, resulting in more difficult meaning construction over an extended time period if the language processor encounters inconsistent incoming information.

The key novel finding, however, concerns the processing of socio-normative materials in the present text comprehension task. That is, morally acceptable compared with morally unacceptable sentences triggered a larger ERP positivity starting at ~320 ms that peaked at ~500 ms. This finding clearly contrasts with the presence of an N400 effect observed for knowledge-based materials. However, it accords with results from previous studies using explicit moral judgment tasks that showed the processing of moral information to be associated with late ERP positivities (Chen *et al.*, 2009; Van Berkum *et al.*, 2009; Sarlo *et al.*, 2012). As in these previous moral decision-making studies, we relate the present ERP positivity to the LPP (or P300) component. This inference is in line with the observation of similar LPP effects triggered by evaluative inconsistencies among stimulus dimensions that are task-irrelevant (e.g. Ito and Cacioppo, 2000) and by evaluatively incongruent prime-target pairs (Herring *et al.*, 2011). Furthermore, ERP studies using text comprehension procedures concerned with both intentional and spontaneous (implicit) trait or goal inferences observed larger LPP (P300) amplitudes in a similar time range when critical words indicated trait- or goal-inconsistent behavior (e.g. Bartholow *et al.*, 2001; Van Duynslaeger *et al.*, 2007; Van der Cruyssen *et al.*, 2009). Therefore, it seems reasonable to conclude that the enhanced ERP positivity (or LPP) to moral transgressions reflects the operation of an implicit evaluative categorization process.

Nevertheless, one might argue that the LPP effect found for moral items is driven by item plausibility given the present finding of slightly lower plausibility ratings for materials containing a socio-normative violation than those that do not. We view this as being an unlikely possibility, however, since for language materials one typically observes a larger N400 for sentences describing (mildly) implausible than plausible situations (Van de Meerendonk *et al.*, 2010; Filik and Leuthold, 2013). This accords with the present observation of larger N400 amplitudes for less plausible knowledge-inconsistent than consistent materials (for plausibility ratings see Supplementary Data). Therefore, we are confident that plausibility cannot explain our LPP findings. Still, future studies should manipulate the plausibility of socio-normative scenarios and of moral transgressions to investigate possible influences on the ERP waveform and their time-course.

Importantly, we believe, for several, reasons that our LPP findings provide strong evidence for the assumption that socio-normative materials trigger an implicit evaluative process. Firstly, knowledge-based violations did not elicit a larger positivity but only an extended N400-like effect, suggesting that the present LPP effect is material specific. Secondly, participants performed a simple reading task rather than an explicit judgment task, as in previous ERP studies of moral decision-making (Chen *et al.*, 2009; Van Berkum *et al.*, 2009; Sarlo *et al.*, 2012). Thirdly, the present moral acceptability effect had an earlier onset (at ~320 ms) compared with LPP effects observed in these previous moral ERP studies. Finally, also in contrast to these studies, socio-normative items were randomly presented together with other (unrelated) knowledge-based and neutral filler materials. Consequently, it seems unlikely that participants were biased toward engaging in deliberate moral evaluations. We therefore hypothesize that during a first step, incoming social information is involuntarily evaluated and categorized as good or bad (e.g. Cunningham and Zelazo, 2007). This hypothesis accords with Greene *et al.*'s (2001; also see Suter and Hertwig, 2011; Conway and Gawronski, 2013) parallel model of moral judgments, at least as far as the involvement and more rapid time-course of emotional compared with cognitive processes is concerned. Moreover, it is in line with Haidt's (2001) SIM, which assumes that moral judgments are primarily based on moral intuitions, that is, an automatic evaluative process which promotes a feeling of good–bad or like–dislike, whereas controlled reasoning processes dominate the construction of post-hoc justifications of the intuition-based judgments.

One interesting issue for the present study concerns the absence of an N400 effect for the socio-normative violations, in contrast to

discourse-based value and emotion violations shown in previous studies. For example, Van Berkum *et al.* (2009) found that value-inconsistent words elicited a slightly larger centroparietal N400 that followed an earlier, more widely distributed ERP positivity. Also, Leuthold *et al.* (2012) and León *et al.* (2010) found that socio-emotional expectancy violations embedded in a discourse context triggered a larger N400 than non-violations. Assuming that our socio-normative violations relate to value-based inconsistencies and also to emotional violations, it is somewhat surprising that we didn't observe an N400 effect. In our view, one might account for the absence of an N400 effect in at least two different ways.

One possibility is that a rather small N400 amplitude effect was indeed triggered by violations of the personal value system, as indicated in Van Berkum *et al.*'s (2009) study. However, a simultaneously occurring and larger moral acceptability effect on a similarly distributed ERP positivity (LPP) might have masked the one in the N400 time period. Another possibility is offered by a recent study of Delaney-Busch and Kuperberg (2013) that examined ERP effects triggered by emotional text materials. In contrast to other studies reporting discourse-based emotional violations eliciting a larger N400 (e.g. León *et al.*, 2010; Leuthold *et al.*, 2012), it was the amplitude of a late centroparietal positivity (500–700 ms) rather than N400 amplitude which was enhanced in emotion-incongruent conditions. Delaney-Busch and Kuperberg interpreted their ERP findings in relation to the affective primacy hypothesis (cf. Storbeck and Clore, 2007), according to which prioritizing of emotional information processing occurs at the expense of cognitive, in this case semantic processing. In addition, they assumed that the occurrence of an N400 effect under such conditions depends on the contextual constraints imposed on the processing of the forthcoming critical word. This assumption might explain the N400 effect observed in the ERP study of Leuthold *et al.* (2012), because quite constraining discourse contexts were employed in this work and, hence, valence-inconsistent words were presumably quite unexpected. In contrast, the demands of meaning construction as reflected by the N400 should become comparable for valence-consistent and inconsistent words if less constraining contexts were used with respect to the valence of the critical word. This interpretation also accords with ERP findings in the evaluative priming paradigm, where the need for meaning integration into the context is negligible, and the LPP but not the N400 was observed to be sensitive to evaluative incongruency (Herring *et al.*, 2011). Certainly, it is an interesting task for future studies to more specifically investigate the role of contextual constraints in the processing of socio-normative materials.

In conclusion, the present research has shed important new light on the processing of socio-normative information under conditions where no explicit evaluative judgments were demanded. Most importantly, whereas materials containing knowledge-based inconsistencies triggered a larger N400, we observed a larger ERP positivity (LPP) starting at 320 ms after the onset of the critical word that indicated a social norm violation. We take this relatively early ERP positivity to reflect an implicit evaluative (good–bad) categorization process that is engaged during the online processing of moral transgressions. Building on the present work, we are confident that future studies will more specifically reveal the nature of implicit evaluative processes by testing how manipulations of contextual and emotional factors modulate the online processing of socio-normative information during text comprehension.

SUPPLEMENTARY DATA

Supplementary data are available at *SCAN* online

Conflict of Interest

None declared.

REFERENCES

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.

Blank, H. (2009). Remembering: a theoretical interface between memory and social psychology. *Social Psychology*, *40*, 164–75.

Bohan, J., Leuthold, H., Hijikata, Y., Sanford, A.J. (2012). Anomalies at the borderline of awareness: an ERP study. *Neuropsychologia*, *50*, 3174–84.

Brainard, D.H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–36.

Cacioppo, J.T., Crites, S.L., Berntson, G.G., Coles, M.G.H. (1993). If attitudes affects how stimuli are processed, should they not affect the event-related brain potential? *Psychological Science*, *4*, 108–12.

Chen, P., Qiu, J., Li, H., Zhang, Q. (2009). Spatiotemporal cortical activation underlying dilemma decision-making: an event-related potential study. *Biological Psychology*, *82*, 111–15.

Conway, P., Gawronski, B. (2013). Deontological versus utilitarian inclinations in moral decision-making: a process dissociation approach. *Journal of Personality and Social Psychology*, *104*, 216–35.

Crites, S.L., Jr, Cacioppo, J.T., Gardner, W.L., Berntson, G.G. (1995). Bioelectrical echoes from evaluative categorizations: II. A late positive brain potential that varies as a function of attitude registration rather than attitude report. *Journal of Personality and Social Psychology*, *68*, 997–1013.

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, and Behavioral Neuroscience*, *13*, 473–90.

DeLong, K.A., Urbach, T.P., Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*, 1117–21.

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.

Dickter, C.L., Bartholow, B.D. (2007). Racial ingroup and outgroup attention biases revealed by event-related brain potentials. *Social Cognitive and Affective Neuroscience*, *2*, 189–98.

Filik, R., Leuthold, H. (2008). Processing local pragmatic anomalies in fictional contexts: evidence from the N400. *Psychophysiology*, *45*, 554–58.

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.

Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M., Cohen, J.D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*, 389–400.

Greene, J.D., Morelli, S.A., Lowenberg, K., Nvstrom, L.E., Cohen, J.D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*, 1144–54.

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*, 2105–8.

Hagoort, P., Hald, L., Bastiaansen, M., Petersson, K.M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*, 438–41.

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814–34.

Haidt, J. (2007). The new synthesis in moral psychology. *Science*, *316*, 998–1002.

Haidt, J., Kesebir, S. (2010). Morality. In: Fiske, S., Gilbert, D., Lindzey, G., editors. *Handbook of Social Psychology* 5th edn. Hoboken, NJ: Wiley, pp. 797–832.

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.

Hillyard, S.A., Münte, T.F. (1984). Selective attention to colour and location: an analysis with event-related brain potentials. *Perception & Psychophysics*, *36*, 185–98.

Ito, T.A., Cacioppo, J.T. (2000). Electrophysiological evidence of implicit and explicit categorization processes. *Journal of Experimental Social Psychology*, *36*, 660–76.

Ito, T.A., Thompson, E., Cacioppo, J.T. (2004). Tracking the timecourse of social perception: the effects of racial cues on event-related brain potentials. *Personality and Social Psychology Bulletin*, *30*, 1267–80.

Kahane, G. (2012). On the wrong track: process and content in moral psychology. *Mind and Language*, *25*, 519–45.

Kissler, J., Assadollahi, R., Herbert, C. (2006). Emotional and semantic networks in visual word processing: insights from ERP studies. *Progress in Brain Research*, *156*, 147–83.

Kleiner, M., Brainard, D., Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception* (ECVP Abstract Supplement)*36*, 14.

Kohlberg, L. (1981). *The Philosophy of Moral Development*. San Francisco, CA: Harper.

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*, 621–47.

Kutas, M., Hillyard, S.A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*, 161–63.

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*, 863–73.

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*, 457–66.

Luck, S.J., Hillyard, S.A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology*, *31*, 291–308.

Nolan, H., Whelan, R., Reilly, R.B. (2010). FASTER: fully automated statistical thresholding for EEG artifact rejection. *Journal of Neuroscience Methods*, *192*, 152–62.

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.

Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*, 437–42.

Sanford, A.J., Leuthold, J., Bohan, J., Sanford, A. (2011). Anomalies at the borderline of awareness: an ERP study. *Journal of Cognitive Neuroscience*, *23*, 514–23.

Sarlo, M., Lotto, L., Manfrinati, A., et al. (2012). Temporal dynamics of cognitive-emotional interplay in moral decision-making. *Journal of Cognitive Neuroscience*, *24*, 1018–29.

Storbeck, J., Clore, G.L. (2007). On the interdependence of cognition and emotion. *Cognition & Emotion*, *21*, 1212–37.

Suter, R.S., Hertwig, R. (2011). Time and moral judgment. *Cognition*, *119*, 454–58.

Tulving, E. (2002). Episodic memory: from mind to brain. *Annual Review of Psychology*, *53*, 1–25.

Van Berkum, J.J.A. (2012). The electrophysiology of discourse and conversation. In: Spivey, M., Joanisse, M., McRae, K., editors. *The Cambridge Handbook of Psycholinguistics*. Cambridge: Cambridge University Press, pp. 589–612.

Van Berkum, J.J.A., Brown, C.M., Zwitserlood, P., Kooijman, V., Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 443–67.

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*, 657–71.

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*, 1092–99.

Van Berkum, J.J.A., Van den Brink, D., Tesink, C.M.J.Y., Kos, M., Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*, 580–91.

Van de Meerendonk, N., Kolk, H.H.J., Vissers, C.T.W.M., Chwilla, D.J. (2010). Monitoring in language perception: mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience*, *22*, 67–82.

Van der Cruyssen, L., Van Duynslaeger, M., Cortoos, A., Van Overwalle, F. (2009). ERP time course and brain areas of spontaneous and intentional goal inferences. *Social Neuroscience*, *4*, 165–84.

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*, 174–88.

Waldmann, M.H., Nagel, J., Wiegmann, A. (2012). Moral judgment. In: Holyoak, K.J., Morrison, R.G., editors. *The Oxford Handbook of Thinking and Reasoning*. Oxford: Oxford University Press, pp. 364–89.

White, K.R., Crites, S.L., Jr, Taylor, J.H., Corral, G. (2009). Wait, what? Assessing stereotype incongruities using the N400 ERP component. *Social, Cognitive, and Affective Neuroscience*, *4*, 191–98.