A JOURNAL OF THE ASSOCIATION FOR PSYCHOLOGICAL SCIENCE aps

Research Article

Right or Wrong?

The Brain's Fast Response to Morally Objectionable Statements

Jos J.A. Van Berkum,[1,2,3] Bregje Holleman,[4] Mante Nieuwland,[1,5] Marte Otten,[1,6] and Jaap Murre[1]

[1]Department of Psychology, University of Amsterdam; [2]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands; [3]Donders Institute, Centre for Cognitive Neuroimaging, Radboud University Nijmegen; [4]Utrecht Institute for Linguistics/OTS, Utrecht University; [5]Department of Psychology, Tufts University; and [6]Department of Psychology, Harvard University

ABSTRACT—*How does the brain respond to statements that clash with a person's value system? We recorded event-related brain potentials while respondents from contrasting political-ethical backgrounds completed an attitude survey on drugs, medical ethics, social conduct, and other issues. Our results show that value-based disagreement is unlocked by language extremely rapidly, within 200 to 250 ms after the first word that indicates a clash with the reader's value system (e.g., "I think euthanasia is an acceptable/unacceptable. . ."). Furthermore, strong disagreement rapidly influences the ongoing analysis of meaning, which indicates that even very early processes in language comprehension are sensitive to a person's value system. Our results testify to rapid reciprocal links between neural systems for language and for valuation.*

People disagree over things of fundamental importance, such as whether euthanasia is acceptable, whether it is okay to joke about somebody's religion, and whether one's country should shut out economic refugees. The moral values behind these disagreements are frequently debated with language and—in attitude surveys—probed through language. We used electroencephalographic (EEG) data to study how rapidly values are brought to bear on processing as people read an attitude-survey statement, what the neural consequences of such value-based processing are, and whether values can in fact influence ongoing linguistic-semantic analysis, the process that builds meaning from a sequence of words.

Address correspondence to Jos J.A. Van Berkum, Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands, e-mail: jos.vanberkum@mpi.nl.

These questions are at the intersection of disciplines that have had little interaction (Holleman & Murre, 2008). Language processing is the subject matter of linguistics and psycholinguistics. Research in these fields tends to capitalize on affectively neutral knowledge of language and the world (*cold cognition*), often using stimuli that people do not really care about. Moral values, attitudes, and emotions have been studied in disciplines that pay more attention to affective valence (*hot cognition*), such as social and clinical psychology and the psychology of emotion. In these fields, stimuli are designed to be motivationally relevant and to recruit affective (emotion) systems. Language is sometimes used as a vehicle, but the studies rarely explore the interaction between language and affect. As a result, little is known about how the neural systems that support linguistic communication are coordinated with those that support morality, valuation, and emotion.

To explore this language-value interface, we asked two groups of Dutch respondents with opposing value systems to complete a realistic attitude survey on societal matters while we recorded their EEG. The first group consisted of members of a relatively strict Christian party, of interest to us because people from this community tend to have relatively stable and outspoken ideas about many morally relevant issues in society. For the second group, we sampled from non-Christians who voted for various political parties that take a diametrically opposed stance on the same issues.[1]

[1]Our particular choice of groups was driven solely by the need to test respondents with relatively predictable and outspoken opposing views on many morally relevant issues. It is perhaps no surprise that a religion-related selection process turned out to be the most practical one. But the precise choice is irrelevant to our current concerns: This study was not aimed at identifying differences between, or similarities across, *specific* religious (or nonreligious) groups.

Psychological Science RESEARCH, THEORY, & APPLICATION IN PSYCHOLOGY AND RELATED SCIENCES

Our interest was in the event-related potential (ERP) response to the first word that indicated a statement clashed with the reader's value system. For each group, we compared ERP responses to value-inconsistent critical words with ERP responses to value-consistent critical words. For example, consider the following two statements:

(1) I think euthanasia is an acceptable course of action.
(2) I think euthanasia is an unacceptable course of action.

For respondents in the strict Christian (SC) group, we compared ERP responses to the value-inconsistent word *acceptable* in (1) with ERP responses to the value-consistent word *unacceptable* in (2). For respondents in the non-Christian (NC) group, we compared ERPs across the same statements, but the comparison was in the opposite direction because the value-inconsistent word was *unacceptable*, and the value-consistent word was *acceptable*.

What happens when people come across a strongly value-inconsistent word? A working assumption in cognitive survey research is that respondents first read an entire statement and then decide how they feel about it (see Tourangeau, Rips, & Rasinski, 2000, for discussion). However, research in other fields suggests that the initial valuation of an attitude-survey statement may occur very rapidly, as the statement unfolds. Psycholinguistic studies have shown that the meaning of a sentence is incrementally computed as it unfolds and that such incremental sense making takes the wider interpretive context (e.g., the identity of the speaker) into account as each word comes in, within only a few hundred milliseconds (for review, see Van Berkum, 2008, in press-a). Such observations make it unlikely that readers delay in bringing their value system to bear on interpretation. If anything, the evolutionary significance of being able to rapidly tell good from bad suggests that valuations might be among the first bits of information to be computed.

In line with this expectation, research on feelings and emotions has revealed that the human brain responds extremely rapidly to positive or negative stimuli, sometimes within a mere 100 to 150 ms (e.g., Pizzagalli et al., 2002; Schupp et al., 2004; Smith, Cacioppo, Larsen, & Chartrand, 2003). Related work in social psychology has shown that relatively stable ideas about whether something—or somebody—is good or bad can have very rapid, implicit effects on processing (see, e.g., Cunningham & Zelazo, 2007; Ito & Cacioppo, 2000; Satpute & Lieberman, 2006). And contrary to the classic rationalist idea that moral judgment is based on careful deliberation, research on moral decision making suggests that such judgment is usually grounded in quick, automatic feelings of approval or disapproval (Greene, 2003; Haidt, 2001). The latter work indicates that it is not just "simple" stimuli (e.g., big hairy spiders, cute-looking babies) that are rapidly valuated; complex moral scenarios can also rapidly engage the valence, or affect, system.

These various findings jointly suggest that a strongly value-inconsistent statement may well engage the affect system very rapidly, at the first word that makes the objectionable contents of the statement apparent. What might be the neural signature of this language-value clash? One relevant ERP component is the *late positive potential*, or *LPP* (Cacioppo, Crites, Berntson, & Coles, 1993), which is elicited by stimuli with emotional content. The LPP typically has its onset somewhere between 300 and 500 ms, lasts for several hundreds of milliseconds, and has a maximum over centro-parietal scalp regions (around electrode Pz). It is elicited by emotional pictures and words alike, can be observed even when participants are not performing an explicit rating task, and varies in amplitude with subjective ratings of emotional arousal. For these and other reasons, the LPP is taken to reflect the affect-induced intensified processing of motivationally important stimuli (e.g., Cacioppo et al., 1993; Cacioppo, Larsen, Smith, & Berntson, 2004; Holt, Lynn, & Kuperberg, in press; Kisley, Wood, & Burrows, 2007; Sabatinelli, Lang, Keil, & Bradley, 2007; Schupp et al., 2000, 2004; Smith et al., 2003).

Of particular importance to our study is the fact that negatively valenced stimuli tend to generate stronger LPP responses than positive ones (e.g., Cacioppo, Crites, Gardner, & Berntson, 1994; Holt et al., in press; Ito, Larsen, Smith, & Cacioppo, 1998; Kisley et al., 2007; Sabatinelli et al., 2007; Smith et al., 2003), an asymmetry taken to reflect a more general *negativity bias* in human cognition. Put simply, the idea is that for survival, it generally pays to rapidly allocate extra attention to potentially aversive stimuli (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Rozin & Royzman, 2001). Although this mechanism did not evolve in the context of language processing, morally objectionable statements do signal something potentially aversive. Hence, if the unfolding message of a statement is evaluated sufficiently rapidly, value-inconsistent critical words might well elicit an LPP effect.

Value-inconsistent words may also increase the amplitude of the *N400* component. The N400 is a negative ERP deflection that begins to develop around 250 ms after a written word, peaks around 400 ms, lasts for several hundreds of milliseconds, and has a scalp distribution roughly similar to that of the LPP (i.e., a centro-parietal maximum, around electrode Pz). In language comprehension, the N400 reflects neural processes involved in relating the meaning of a word to its context, and a larger amplitude (i.e., a more negative N400) indexes more difficult or intensified processing (e.g., as in the case of "He took his coffee with cream and dog"; Kutas & Hillyard, 1980, 1984; see Kutas, Van Petten, & Kluender, 2006, for a review). Words that render a statement inconsistent with personal values may well be unexpected, emotionally salient, or otherwise attention grabbing. As a result, they might call for intensified processing of meaning, and therefore elevate the N400.

METHOD

Respondents

The 21 respondents in the SC group (19 right-handed, 2 left-handed; mean age = 46 years, range = 31–62) were members of

a relatively strict Christian party, the Dutch *Staatkundig Gereformeerde Partij* (SGP, or Reformed Political Party). The 22 respondents in the NC group (18 right-handed, 4 left-handed; mean age = 45 years, range = 30–62; all self-designated as "nonreligious") voted for political parties with moral-ethical programs opposite to that of the SGP. Because at the time of testing only men were allowed to join the SGP, all participants were male. Groups were matched on educational background and verbal working memory as measured by the Dutch Reading Span Test (Van den Noort, Bosch, Haverkort, & Hugdahl, 2008; SC group: mean = 68.1, range = 48–91; NC group: mean = 71.7, range = 49–90).

Materials

Working from party programs, we constructed 158 statements that SC and NC respondents would be expected to disagree over. In each statement, a value object was followed by a critical evaluative word that made the core message of the statement sufficiently clear (e.g., "I think *euthanasia* is an *unacceptable* course of action"; italics added for expository purposes). Each statement had an SC-consistent and an NC-consistent variant (see Table 1), differing only in the critical evaluative word. Most statements (80%) contained self-referential terms (e.g., "I think. . . ").

In a Web survey, 150 SC respondents and 150 NC respondents, again all male, rated their agreement or disagreement with these 158 statements using a 5-point scale. We then computed, for each statement, the mean inconsistency effect across variants for SC respondents, across variants for NC respondents, across groups for the SC-consistent variant, and across groups for the NC-consistent variant. For all 90 statements selected for the main experiment, all of these comparisons yielded a difference of at least 0.5 point in the intended direction. Across these selected statements, the mean inconsistency effects were 2.8, 2.3, 2.5, and 2.5 points for the four comparisons, respectively. Thus, the average response shift between statement variants, as well as between respondent groups, was equal to about half the rating scale, and in the expected direction. Furthermore, the 90 critical words in the SC- and NC-consistent variants were matched on mean presentation duration (554 vs. 550 ms), length (9.7 vs. 9.2 letters), and Celex word frequency (51.5 vs. 46.6 per million, 2.6 vs. 2.7 log-transformed). (The complete list of statements can be obtained at www.josvanberkum.nl.)

TABLE 1
Translated Examples of Statements Used in the Experiment

I think euthanasia is an *unacceptable/acceptable* course of action.
Watching TV to relax is *wrong/fine* in my opinion.
I think the increasing emancipation of women is a *negative/positive* development.
A society that condones abortion is a *bad/good* society.
If my child were homosexual, I'd find this *hard/easy* to accept.
The use of soft drugs should be *forbidden/allowed* in my opinion.
In a bad marriage, divorce is an *unacceptable/acceptable* solution.

Note. Critical words are in italics; in each sentence, the first critical word is consistent with the values of the strict-Christian group, and the second is consistent with the values of the non-Christian group. Each respondent saw just one version of each statement.

If value-inconsistent statements rapidly engage the affect system, the two respondent groups, given their differing value systems, would be expected to show the same ERP effects, but for opposite variants of the critical statements. However, as an additional check on the validity of our research design and the associated assumptions, we included the same 90 critical evaluative words in 90 control statements, in which these words were mentioned before the issue to be evaluated (e.g., "I think it is *acceptable/unacceptable* that people consider euthanasia"; italics added). ERP responses to the critical words in these control statements were not expected to differ as a function of the reader's value system because at this early point in the statements the reader had not yet seen the particular issue to be evaluated (e.g., "that people consider euthanasia").

In the ERP study, the two types of statements were pseudo-randomly mixed and, as is customary in survey research, presented in thematically coherent blocks (block order was counterbalanced, and each block contained as many SC-consistent as NC-consistent statements). We used two different randomizations, as well as a reversed version of each. Identical critical words were separated by at least 12 other statements, and no more than 3 consecutive statements of the same type (critical or control) were allowed. Each respondent saw only one version of each statement and completed a short practice block prior to the experimental blocks.

Procedure

Respondents indicated their agreement or disagreement with 180 statements on a 4-point scale. The direction of the scale was counterbalanced with handedness. Each trial began with a 1,000-ms fixation cross and then a 1,000-ms blank screen before the statement was presented visually word by word. Next, following a 400-ms blank screen, the response-scale display (maximum of 7 s) cued respondents to give their opinion. Instructions emphasized providing an accurate response, and respondents were allowed to skip a response (1.4% of critical trials). Except for statement-final words (always presented for 1,000 ms), each word was presented for 290 ms plus 30 ms per letter, up to a maximum of 590 ms; words were separated by a blank 150-ms interval. Mean duration of critical (and adjacent) words was matched across conditions. The average recording session lasted 50 min, and procedures were approved by the University of Amsterdam Psychology Department ethics committee.

EEG Recording and Analysis

EEG was recorded from 32 Ag/AgCl electrodes (< 10 kΩ) referenced to the left mastoid, amplified with BrainAmps DC am-

plifiers (Brain Products, Gilching, Germany; 500-Hz sampling, 0.03- to 100-Hz band pass), and rereferenced off-line to the mastoid average. After removal of eye artifacts with Independent Components Analysis (Jung et al., 2000), the data were segmented in epochs from 500 ms before to 1,200 ms after the onset of the critical evaluative word and baseline-corrected using the 150 ms preceding that onset.[2] Signals that exceeded ±75 μV or that had a linear drift (beginning before the critical word) of at least ±40 μV were rejected as artifacts (4.1% of trials).

To increase power, we analyzed critical statements only if they had attracted a strong group-compatible response, that is, a "strongly agree" or "strongly disagree" response in line with the average value system of the group the respondent belonged to (as assessed for each statement in the Web pretest). Thus, group-incompatible responses (13.3%), moderate responses (18.3%), and skipped responses (1.4%) were excluded. Average response times were comparable across groups and responses—SC group: 1,080 ms for "agree" responses, 1,052 ms for "disagree" responses; NC group: 1,072 ms for "agree" responses, 1,061 ms for "disagree" responses. In the case of control statements, only those that were not responded to (2.0%) were removed. Remaining EEG epochs were averaged per participant, statement type, and response ("agree" or "disagree"), and mean amplitude values were analyzed with repeated measures analysis of variance, using Greenhouse-Geisser/Box's epsilon correction for F tests with 2 or more degrees of freedom.

RESULTS

As illustrated in Figure 1, value-inconsistent words elicited a small N400 effect (−0.50 μV) with a centro-parietal maximum around 400 ms after word onset (375–425 ms) in a six-electrode region around Pz, $F(1, 41) = 5.11$, $MSE = 6.38$, $p = .029$, $p_{rep} = .94$, $\eta_p^2 = .11$.[3] Value-inconsistent words also elicited positivities around 200 to 250 ms (+0.50 μV), $F(1, 41) = 7.54$, $MSE = 21.13$, $p = .009$, $p_{rep} = .97$, $\eta_p^2 = .16$, and around 500 to 650 ms (+0.46 μV), $F(1, 41) = 8.74$, $MSE = 15.55$, $p = .005$, $p_{rep} = .98$, $\eta_p^2 = .18$ (results across all electrodes). The same triphasic pattern of ERP deflections was observed for the two groups ($F < 1$ for all interactions with group), on exactly opposite statements.[4]

As we argue later, we take the late positivity to be an LPP effect. However, its scalp distribution is not the canonical one: Instead of being largest over centro-parietal scalp sites, it actually has a circular attenuation there (the yellow roundish area over the back of the head in Fig. 1a). The fact that the scalp region where this attenuation occurred is virtually identical to the region showing the small N400 effect suggests that the N400 effect at least partly overlapped with a longer-lasting (and more broadly distributed) LPP effect.

The data from control statements (Fig. 2) confirmed that the observed effects hinged on values interacting with statement content: When the same critical words were presented in a neutral prior sentence context (e.g., "I think it is *acceptable*. . . "), no differential ERP response was observed—200–250 ms: $F(1, 41) = 0.69$, $MSE = 14.27$, $p = .410$, $p_{rep} = .72$, $\eta_p^2 = .02$ (across all electrodes); 375–425 ms: $F(1, 41) = 1.35$, $MSE = 5.92$, $p = .252$, $p_{rep} = .79$, $\eta_p^2 = .03$ (across the six posterior electrodes); 500–650 ms: $F(1, 41) = 0.31$, $MSE = 12.44$, $p = .580$, $p_{rep} = .65$, $\eta_p^2 = .01$ (across all electrodes). These observations held for both groups ($F < 1$ for all interactions with group).

As suggested by a comparison between Figures 1 and 2, from about 350 to 400 ms onward, the ERPs were on the whole more positive for critical words than for control words, regardless of value inconsistency. This centro-parietally maximal positivity might be a general LPP to value-relevant words (regardless of their specific valuation) or some other correlate of explicit evaluation-associated processing. However, critical and control words were not matched on important variables, including ordinal and syntactic sentence position. We therefore focus on findings associated with the predetermined, well-controlled aspects of the design.

DISCUSSION

When people fill out a realistic attitude survey, the first word indicating that a statement clashes with the reader's value system elicits a very rapid and characteristic neural response: an early and broadly distributed positivity between 200 and 250 ms, a small but standard N400 effect peaking at 400 ms, and a broadly distributed late positivity around 500 to 650 ms. Furthermore, whereas, say, "I think euthanasia is an acceptable. . ." elicits this response in individuals with a strict Christian value system, the opposite statement ("I think euthanasia is an unacceptable. . . ") elicits the same triphasic response in individuals with an opposing value system. These midsentence responses reveal that people valuate what they read incrementally, on a word-by-word basis. In addition, each of the three ERP effects tells its own story about the language-value interface.

Personal Values Affect Early Sense Making

The amplitude of the N400 is generally taken to index the difficulty of early sense making (or retrieval of conceptual memory

[2]The preceding value object was typically distributed across several words (e.g., "the increasing emancipation of women"), which prohibited a sensible ERP analysis of the value objects.

[3]In a topography-oriented analysis with factors of anterior/posterior location and left/right hemisphere, the value-inconsistency effect in the 375- to 425-ms latency range was significantly larger over the posterior than the anterior area, $F(1, 41) = 4.22$, $MSE = 0.69$, $p = .046$, $p_{rep} = .92$, $\eta_p^2 = .09$.

[4]There were too few moderate responses (18.3%) to support a separate ERP analysis, but when we included these responses in an analysis with the strong responses, all three effects were attenuated (from 0.50 μV to 0.32 μV, from −0.50 μV to −0.32 μV, and from 0.46 μV to 0.34 μV). Because mild disagreement should have more limited consequences than severe disagreement, this finding supports our interpretation.

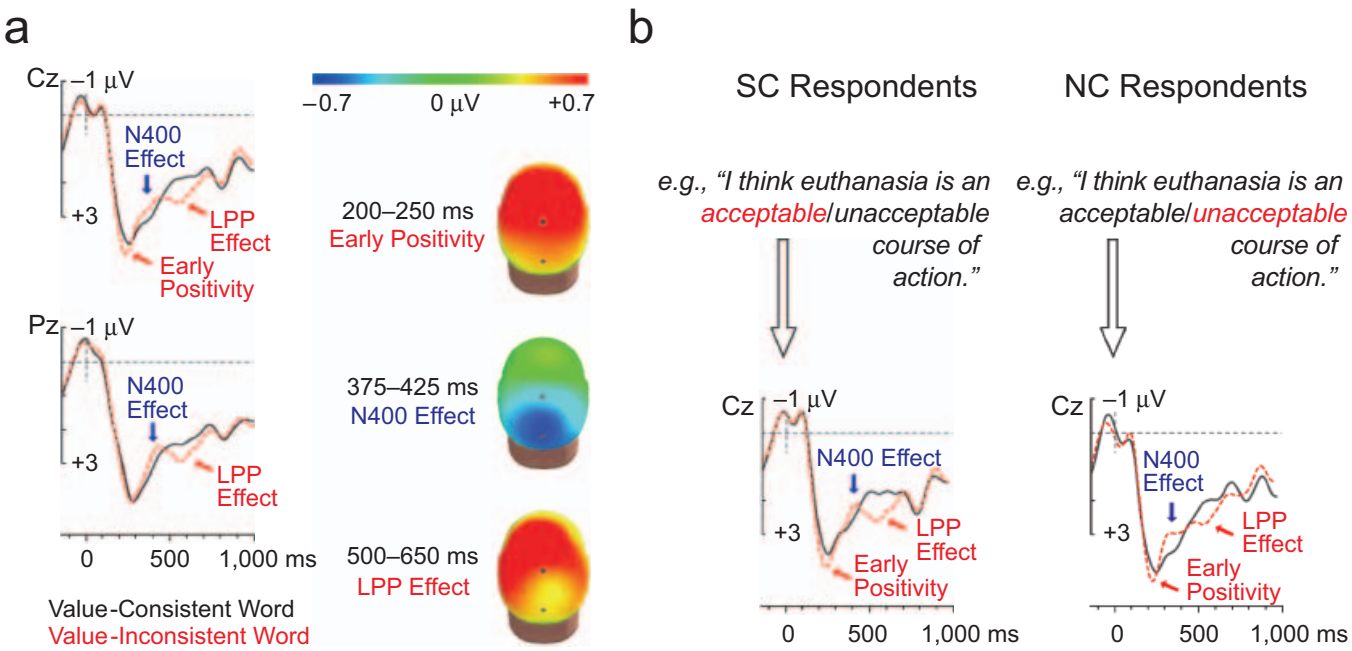

Fig. 1. The brain signature of strong disagreement during reading of attitude-survey statements. The waveforms in (a) show the grand-average event-related potentials (ERPs) to value-consistent and value-inconsistent critical words, pooled across the strict-Christian (SC) and non-Christian (NC) groups of respondents. Shown next to the waveforms are the associated scalp distributions of the three differential effects: an early positivity, the N400, and the late positive potential (LPP). The waveforms in (b) show the ERPs for the SC and NC groups separately. Grand-average ERP signals were filtered (5-Hz low pass, 24-dB slope) for expository purposes, voltage is displayed with negative up, and time is shown relative to the onset of the critical word. Scalp-distribution maps are spline-interpolated isovoltage maps of the grand-average ERP difference between value-inconsistent and value-consistent words (inconsistent minus consistent) in specific latency ranges; the distributions are rendered on a three-dimensional head model with Cz and Pz marked.

in the service thereof; see Kutas et al., 2006; Lau, Phillips, & Poeppel, 2008; Van Berkum, in press-b). The small N400 effect for value-inconsistent words therefore suggests that people briefly experience difficulty making sense of an unfolding statement that strongly clashes with their personal values. Why might this happen? Given that little research has shed light on the language-value interface so far, several causal scenarios need to be considered. First, it is possible that sentence fragments like "I think euthanasia is an . . ." lead one to expect particular words or concepts, which depend on one's specific values. Such expectations could be based on implicit value-mediated priming triggered by particular core concepts (e.g., "euthanasia," "emancipation of women"; Morris, Squires, Taber, & Lodge, 2003) or on the precise message conveyed by the portion of the statement that has been read (as the incrementally computed exact message is known to support specific lexical predictions; Otten & Van Berkum, 2007, 2008). Either way, to the extent that value-based expectations render the critical word less expected, it will elicit a larger N400 response (e.g., Federmeier, 2007; Kutas & Hillyard, 1984; Otten & Van Berkum, 2007).

Second, if a coarse valuation of a sentence fragment like "I think euthanasia is an acceptable . . . " becomes available 200 to 250 ms after onset of the word *acceptable*, the affective salience of the statement at that point might actually lead to an enhanced semantic analysis of that word. Support for this possibility comes from reports that N400 components increase in response to emotionally salient words (e.g., *criminal*) in neutral, unpredictive contexts (Holt et al., in press; see also Bernat, Bunce, & Shevrin, 2001). The temporal extent of word-elicited N400 responses itself indicates that sense making takes time, so there would be time for rapidly delivered intermediate results to feed back into the same analysis.

Third, and more radically, it is possible that the observed N400 indexes processing difficulties in *initial* meaning construction. Partly because of a historic focus on context-free, timeless sentence meaning, language researchers tend to disregard valence as a semantic primitive. However, valence is rooted in relevance to survival and well-being (Cacioppo et al., 2004; Damasio, 2004), which makes it a plausible core ingredient of meaning. If the valence of a concept is stored as part of its meaning for a given person (cf. Morris et al., 2003), the affective valuation of an unfolding statement becomes an integral part of computing statement meaning. Valence-based conceptual mismatches would then be on a par with standard semantic anomalies (e.g., "He took his coffee with cream and dog"), and should generate an N400 effect for the same reasons.

Our N400 findings have led us to consider several (not necessarily mutually exclusive) accounts of how strong value-based disagreement might interfere with initial sense making. Whether such an impact is limited to explicitly value-probing contexts is currently unknown. But note that our respondents knew

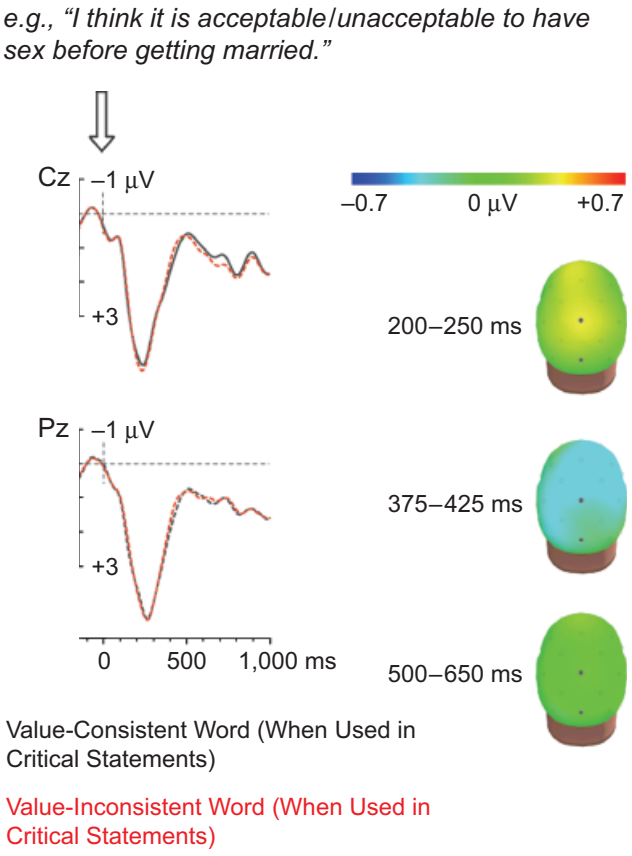

Fig. 2. Results for the control statements: event-related potentials (ERPs) to the same pairs of evaluative words that were used in the critical statements (e.g., "acceptable/unacceptable"), but instead positioned very early in the control statements so that they preceded the value object. Waveforms for words that were value consistent in the critical statements are shown in black, and waveforms for words that were value inconsistent in the critical statements are shown in red. Also shown are the scalp distributions of the associated ERP differences in latency ranges for which the three ERP effects were observed with the critical statements. ERPs and scalp distributions are pooled across the strict-Christian (SC) and non-Christian (NC) groups of respondents. See the caption of Figure 1 for technical details regarding the ERP and scalp-distribution displays.

(at least after the practice block) they would frequently see statements they would not agree with. Hence, whatever the exact mechanism, our N400 findings can be taken to reflect a relatively noncontrolled, automatic influence of valuation on language understanding.

Value-Based Disagreement Rapidly Engages the Affect System

The LPP at 500 to 650 ms after a value-inconsistent word suggests that strongly disagreeable statements automatically recruit additional processing resources, just as negatively valenced single words or pictures do. That is, in line with the more general negativity bias in human cognition (Baumeister et al., 2001; Rozin & Royzman, 2001), the effect indicates that our respondents were taking strongly value-inconsistent statements as potentially aversive stimuli that warranted extra attention. This late positivity is unlikely to be just a decision-related ERP effect, because when readers explicitly decide on the correctness of self-referential statements without a moral-emotional component (e.g., "I go to bed late"), no late positivity emerges (Fischler, Bloom, Childers, Arroyo, & Perry, 1984). Furthermore, if people encounter an emotionally salient word in text without having to make any decision at all, negatively valenced words still elicit an LPP effect relative to positively valenced words (e.g., *criminal* vs. *millionaire*; Holt et al., in press). These findings support our interpretation that the late positivity we observed is an affect-related LPP effect.

Value-inconsistent critical words also elicited a much earlier positivity between 200 and 250 ms. We had not anticipated this effect, and can only speculate about its functional interpretation. Very early neural responses have been reported for other emotional stimuli too (e.g., Kisley et al., 2007; Pizzagalli et al., 2002; Schupp et al., 2004; Smith et al., 2003), but none of these effects resemble the one we observed. For the time being, the most parsimonious interpretation, therefore, is that this earlier positivity is actually the onset of a single long-lasting LPP effect that is briefly canceled out by the opposite-polarity N400 effect. Although LPP effects typically begin to develop somewhat later than 200 ms, very early LPP onsets do occur (e.g., Crites, Cacioppo, Gardner, & Berntson, 1995, Fig. 2, left panel; Ito et al., 1998, Fig. 1). Furthermore, several other researchers have obtained word-elicited ERP effects indicative of an early-onset LPP that is momentarily canceled out by an N400 effect (e.g., see Cacioppo et al., 1993, Fig. 1; Cacioppo et al, 1994, Fig. 2; Crites et al., 1995, Fig. 3, right panel). We are currently examining this possibility by means of magnetoencephalography.

Conclusion

At least three issues for further research remain. First, as in the case of any realistic attitude survey, our task necessarily involved explicit evaluation. Whether comparable results would be obtained in less explicitly evaluative settings is unknown. However, it is worth emphasizing here that in our experiment—and in contrast to most other neurocognition studies—language was used in a natural way, to communicate ideas that are relevant to the situation and the goals at hand.

Second, it is as yet unclear whether the observed neural signature of value-based disagreement directly reflects the unlocking of "deep" moral values (e.g., respect for all that lives). Our ERP findings might hinge on negative connotations at a somewhat more superficial level, involving attitudes toward the issue (e.g., bio-industry) or perhaps the out-group associated with it (e.g., people who do not care about animal suffering). Whether these distinctions matter in actual processing remains to be seen.

Third, our findings were obtained with men only. We have no reason to suspect that the basic mechanisms would be qualitatively different with female participants. However, effect sizes

might differ (see Van den Brink et al., 2009, for socially conditioned N400 effects that are larger for women than for men; see Schupp et al., 1996, for LPP effects that vary with how sex and the specific stimulus interact). Only follow-up research with sex-balanced respondent groups can illuminate this issue.

Our findings testify to the presence of very rapid reciprocal links between neural systems for language and for valuation. Furthermore, the speed and nature of the observed effects raise the possibility that valuation might be an integral part of early language interpretation. Finally, our work shows that it is possible to study the language-value interface by bringing a real "arena of language use" (Clark, 1996) into the cognitive neuroscience lab. After all, although the human species did not evolve in an environment that included attitude surveys, filling out such surveys is a real-world task that uses language for a purpose.

Acknowledgments—We thank Steve Janssen, Lindy Odijk, Rosa Walraven, Linda Zoon, Petra van Alphen, José Kerstholt, Martin Pickering, three reviewers, and the Staatkundig Gereformeerde Partij for their help.

REFERENCES

Baumeister, R.F., Bratslavsky, E., Finkenauer, C., & Vohs, K.D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*, 323–370.

Bernat, E., Bunce, S., & Shevrin, H. (2001). Event-related brain potentials differentiate positive and negative mood adjectives during both supraliminal and subliminal visual processing. *International Journal of Psychophysiology*, *42*, 11–34.

Cacioppo, J.T., Crites, S.L., Berntson, G.G., & Coles, M.G.H. (1993). If attitudes affect how stimuli are processed, should they not affect the event-related brain potential? *Psychological Science*, *4*, 108–112.

Cacioppo, J.T., Crites, S.L., Gardner, W.L., & Berntson, G.G. (1994). Bioelectrical echoes from evaluative categorizations: I. A late positive brain potential that varies as a function of trait negativity and extremity. *Journal of Personality and Social Psychology*, *67*, 115–125.

Cacioppo, J.T., Larsen, J.T., Smith, N.K., & Berntson, G.G. (2004). The affect system: What lurks below the surface of feelings? In A.S.R. Manstead, N.H. Frijda, & A.H. Fischer (Eds.), *Feelings and emotions: The Amsterdam conference* (pp. 223–242). New York: Cambridge University Press.

Clark, H.H. (1996). *Using language*. Cambridge, England: Cambridge University Press.

Crites, S.L., Jr., Cacioppo, J.T., Gardner, W.L., & Berntson, G.G. (1995). Bioelectrical echoes from evaluative categorization: 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., & Zelazo, P.D. (2007). Attitudes and evaluation: A social cognitive neuroscience perspective. *Trends in Cognitive Sciences*, *11*, 97–104.

Damasio, A.R. (2004). Emotions and feelings: A neurobiological perspective. In A.S.R. Manstead, N.H. Frijda, & A.H. Fischer (Eds.), *Feelings and emotions: The Amsterdam conference* (pp. 49–57). New York: Cambridge University Press.

Federmeier, K.D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*, 491–505.

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.

Greene, J.D. (2003). From neural 'is' to moral 'ought': What are the moral implications of neuroscientific moral psychology? *Nature Reviews Neuroscience*, *4*, 847–850.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814–834.

Holleman, B.C., & Murre, J.M.J. (2008). Getting from neuron to checkmark: Models and methods in cognitive survey research. *Applied Cognitive Psychology*, *22*, 709–732.

Holt, D.J., Lynn, S.K., & Kuperberg, G.R. (in press). Neurophysiological correlates of comprehending emotional meaning in context. *Journal of Cognitive Neuroscience*.

Ito, T.A., & Cacioppo, J.T. (2000). Electrophysiological evidence of implicit and explicit categorization processes. *Journal of Experimental Social Psychology*, *36*, 660–676.

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.

Jung, T.P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T.J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology*, *111*, 1745–1758.

Kisley, M.A., Wood, S., & Burrows, C.L. (2007). Looking at the sunny side of life: Age-related change in an event-related potential measure of the negativity bias. *Psychological Science*, *18*, 838–843.

Kutas, M., & Hillyard, S.A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203–205.

Kutas, M., & Hillyard, S.A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*, 161–163.

Kutas, M., Van Petten, C., & Kluender, R. (2006). Psycholinguistics electrified II (1994–2005). In M. Traxler & M.A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 659–724). New York: Elsevier.

Lau, E.F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, *9*, 920–933.

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.

Otten, M., & Van Berkum, J.J.A. (2007). What makes a discourse constraining? Comparing the effects of discourse message and scenario fit on the discourse-dependent N400 effect. *Brain Research*, *1153*, 166–177.

Otten, M., & Van Berkum, J.J.A. (2008). Discourse-based anticipation during language processing: Prediction or priming? *Discourse Processes*, *45*, 464–496.

Pizzagalli, D.A., Lehmann, D., Hendrick, A.M., Regard, M., Pascual-Marqui, R.D., & Davidson, R.J. (2002). Affective judgments of

faces modulate early activity (~160 ms) within the fusiform gyri. *NeuroImage*, *16*, 663–677.

Rozin, P., & Royzman, E. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*, 296–320.

Sabatinelli, D., Lang, P.J., Keil, A., & Bradley, M.M. (2007). Emotional perception: Correlation of functional MRI and event-related potentials. *Cerebral Cortex*, *17*, 1085–1091.

Satpute, A.B., & Lieberman, M.D. (2006). Integrating automatic and controlled processing into neurocognitive models of social cognition. *Brain Research*, *1079*, 86–97.

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

Schupp, H.T., Cuthbert, B.N., Hillman, C., Raymann, R., Bradley, M.M., & Lang, P.J. (1996). ERPs and blinks: Sex differences in response to erotic and violent picture content. *Psychophysiology*, *33*, 75.

Schupp, H.T., Öhman, A., Junghöfer, M., Weike, A.I., Stockburger, J., & Hamm, A.O. (2004). The facilitated processing of threatening faces: An ERP analysis. *Emotion*, *4*, 189–200.

Smith, N.K., Cacioppo, J.T., Larsen, J.T., & Chartrand, T.L. (2003). May I have your attention, please: Electrocortical responses to positive and negative stimuli. *Neuropsychologia*, *41*, 171–183.

Tourangeau, R., Rips, L.J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, England: Cambridge University Press.

Van Berkum, J.J.A. (2008). Understanding sentences in context: What brain waves can tell us. *Current Directions in Psychological Science*, *17*, 376–380.

Van Berkum, J.J.A. (in press-a). The electrophysiology of discourse and conversation. In M. Spivey, K. McRae, & M. Joanisse (Eds.), *The Cambridge handbook of psycholinguistics*. Cambridge, England: Cambridge University Press.

Van Berkum, J.J.A. (in press-b). The neuropragmatics of 'simple' utterance comprehension: An ERP review. In U. Sauerland & K. Yatsushiro (Eds.), *Semantics and pragmatics: From experiment to theory*. New York: Palgrave.

Van den Brink, D., Van Berkum, J.J.A., Bastiaansen, M.C.M., Tesink, C.M.J.Y., Kos, M., Buitelaar, J.K., & Hagoort, P. (2009). *Empathy matters: ERP and brain oscillatory evidence for inter-individual differences in social language processing*. Manuscript submitted for publication.

Van den Noort, M.W.M.L., Bosch, M.P.C., Haverkort, M., & Hugdahl, K. (2008). A standard computerized version of the Reading Span Test in different languages. *European Journal of Psychological Assessment*, *24*, 35–42.

(RECEIVED 9/7/08; REVISION ACCEPTED 1/9/09)