

# Leipzig Affective Norms for German: A reliability study

PHILIPP KANSKE

*Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany  
and Central Institute of Mental Health, Mannheim, Germany*

AND

SONJA A. KOTZ

*Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*

To facilitate investigations of verbal emotional processing, we introduce the Leipzig Affective Norms for German (LANG), a list of 1,000 German nouns that have been rated for emotional valence, arousal, and concreteness. A critical factor regarding the quality of normative word data is their reliability. We therefore acquired ratings from a sample that was tested twice, with an interval of 2 years, to calculate test–retest reliability. Furthermore, we recruited a second sample to test reliability across independent samples. The results show (1) the typical quadratic relation of valence and arousal, replicating previous data, (2) very high test–retest reliability ( $>.95$ ), and (3) high correlations between the two samples ( $>.85$ ). Because the range of ratings was also very high, we provide a comprehensive set of words with reliable affective norms, which makes it possible to select highly controlled subsamples varying in emotional status. The database is available as a supplement for this article at <http://brm.psychonomic-journals.org/content/supplemental>.

The emotional status of words strongly influences their processing, as has been shown in a variety of different tasks including lexical decision (Eviatar & Zaidel, 1991; Kanske & Kotz, 2007; Nakic, Smith, Busis, Vythilingam, & Blair, 2006; Ortigue et al., 2004; Scott, O'Donnell, Leuthold, & Sereno, 2009; Wentura, 2000), memory tasks (Kuchinke et al., 2006; Sim & Martinez, 2005), versions of the Stroop task (van Hooff, Dietz, Sharma, & Bowman, 2008), mental imagery (Osaka, Osaka, Morishita, Kondo, & Fukuyama, 2004), the attentional blink (Mathewson, Arnell, & Mansfield, 2008), attentional orienting (Stormark, Nordby, & Hugdahl, 1995), and emotionality judgments (Maddock, Garrett, & Buonocore, 2003). Using word stimuli in studies of emotional processing offers a number of advantages in comparison with images or videos, because words can be tightly controlled for physical attributes (size, complexity, color composition, luminance), frequency of occurrence in everyday life, or concreteness of the underlying concept. However, in order to exploit these advantages, normative data for word stimuli are needed.

Currently, a number of databases offer affective norms for words in different languages, including English (Altarriba, Bauer, & Benvenuto, 1999; Bradley & Lang, 1999; Stevenson, Mikels, & James, 2007), German (Lahl, Göritz, Pietrowsky, & Rosenberg, 2009; Vö et al., 2009; Vö, Jacobs, & Conrad, 2006), Spanish (Redondo, Fraga,

Padrón, & Comesaña, 2007), and Finnish (Eilola & Havelka, 2010). The applicability of these norms in experimental studies critically depends on the reliability of the measured norms. Only when reliability is high will it be appropriate to select words on the basis of normative data acquired in different participant samples. The present study addresses this issue. We obtained ratings from one sample of participants at two different time points with a time lag of 2 years (T1 and T2) to calculate test–retest reliability. To our knowledge, this is the first report of retest reliability for affective word norms. Furthermore, a second independent group of participants was invited at T2 to calculate reliability of the ratings across samples. Participants rated 1,000 German words for valence, arousal, and concreteness, which have been shown to be the most unambiguous factors explaining variance in word ratings across different dimensions (for details, see Hager & Hasselhorn, 1994; Lahl et al., 2009). To reduce variance on the basis of word class (Osterhout, 1997; Perani et al., 1999), we only included nouns.

Even though it is intuitively advisable for experimental studies to utilize emotional stimuli from a database and to rerate them in the present experimental sample, very few studies do this or even report data on different rating results. There is some indication that the mean ratings of positive, negative, and neutral stimuli in valence and arousal may vary in normative data and reratings (for

---

P. Kanske, [philipp.kanske@zi-mannheim.de](mailto:philipp.kanske@zi-mannheim.de)

---



example, ratings of emotional images: Dillon, Ritchey, Johnson, & LaBar, 2007). Such differences may be due to inherently low reliability of affective norms, or to artifacts stemming from differences in rating instructions, the presentation of stimuli, or response options. We therefore kept these factors constant across all testing time points and provide detailed information on the ratings for use in future studies involving the present database and retesting norms in their samples.

To conclude, the present study probes the reliability of affective norms in German words to provide researchers with a comprehensive test inventory.

## METHOD

### Participants

Two separate samples of 32 native German speakers were recruited in Leipzig in 2006 (Sample 1) and 2008 (Sample 2). Sample 1 included 16 female participants; the mean age was 23.2 years ( $SD = 2.8$ ). Because handedness does not influence the behavioral performance in emotion or concreteness rating tasks, handedness was not controlled for (Rodway, Wright, & Hardie, 2003). The resulting mean laterality quotient (LQ), according to the Edinburgh Handedness Inventory (Oldfield, 1971), was 72.3 ( $SD = 47.8$ ). Participants in Sample 2 (16 female, mean age = 22.5 years,  $SD = 3.1$ ) had a mean LQ of 76.5 ( $SD = 41.5$ ). Sample 1 was tested twice, in 2006 (T1) and in 2008 (T2; mean difference 28.3 months). Twenty-two participants from the original Sample 1 could be reached for a retest (mean age at T1 = 23.3 years,  $SD = 3.1$ , 9 females). All participants reported normal or corrected-to-normal vision. For the database norms, the ratings of all 64 participants were averaged (for Sample 1, the first ratings at T1 were used to exclude a potential interference of repeated measures).

### Materials and Procedure

One-thousand German nouns were selected from a previously rated word list (Kanske & Kotz, 2007) and from the Duden dictionary (*Duden: Die deutsche Rechtschreibung*, 2000). Only one- and two-syllable words were included. The number of letters varied between three and eight. Frequency of usage was taken from the Wortschatz Lexikon of the University of Leipzig (<http://wortschatz.uni-leipzig.de/>) and ranged from 8 to 18 ( $M = 12.6$ ,  $SD = 2.4$ ). Compound nouns were excluded.

For each measurement, participants came to the laboratory for two sessions, during which they rated the words for valence (negative, neutral, positive), arousal (high arousing, low arousing), and concreteness (concrete, abstract). The order of the tasks was counterbalanced. Two ratings were always completed in the first session, and one rating in the second session. Ratings were done on 9-point scales. For valence and arousal ratings, the Self-Assessment Manikins (Bradley & Lang, 1994; Hodes, Cook, & Lang, 1985) were used. For the concreteness rating, the endpoints of the scale (concrete–abstract) were presented as words. We include the exact rating instructions in the online supplement. The assignment of the scale endpoints to the left and right was counterbalanced across participants. Words were presented in uppercase in the center of the screen for a maximum of 5,000 msec. They extended to a maximum of 2.3° of visual angle horizontally and 0.4° vertically from fixation. Presentation of a word ended as soon as the participant pressed a button.

### Description of the Leipzig Affective Norms for German (LANG) Database

The database contains 1,000 German nouns with normative data on emotional valence, arousal, and concreteness, which were rated by 64 participants, including separate ratings from female and male

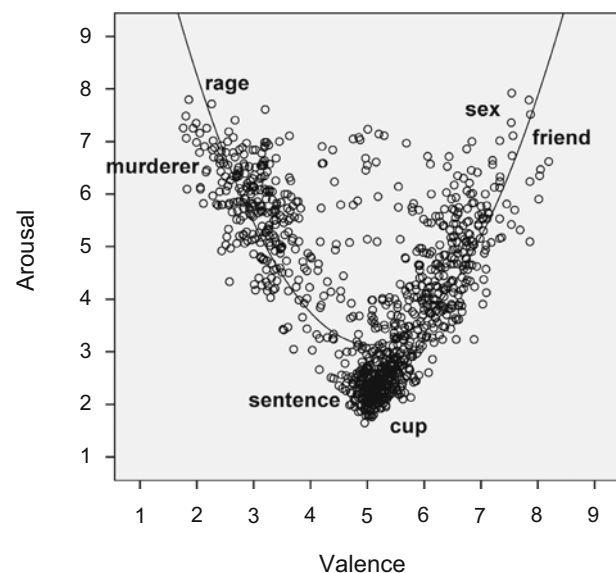
**Table 1**  
Descriptives of the Database

	<i>M</i>	<i>SD</i>	Range
Valence	5.0	2.1	1–9
Arousal	4.2	2.7	1–9
Concreteness	4.1	3.0	1–9
Frequency	12.6	2.4	8–18
Number of letters	5.7	1.2	3–8
Number of syllables	1.8	0.4	1–2

participants. Furthermore, number of letters, number of syllables, and frequency of usage, taken from the Wortschatz Lexikon of the University of Leipzig (<http://wortschatz.uni-leipzig.de/>) are reported. Table 1 shows some descriptive statistics of the database. The database itself is included in the online supplement for this article.

## RESULTS

Figure 1 shows the arousal and valence ratings for each word averaged across all 64 participants. We observed a quadratic relationship ( $r_{\text{quad}} = .77$ ,  $p < .001$ ), confirming the typical distribution of valence and arousal values. Figure 2 shows the valence, arousal, and concreteness ratings of Sample 1 and Sample 2 (A–C) and of Sample 1 at T1 and T2 (D–F), for which we observed linear relations. The respective correlations are reported in Table 2. Test–retest reliability was very high, ranging from  $r = .96$ –.98. The correlations of the ratings in Sample 1 and Sample 2 were also well above  $r = .90$ . Calculations of the correlations for female and male participants separately were slightly lower but still high, and ranged from  $r = .87$  for valence ratings in female participants to  $r = .95$  for the concreteness ratings. All correlations were significant ( $p < .001$ ).



**Figure 1.** Mean valence and arousal ratings for each word across all 64 participants. Example words for neutral, negative, and positive words are given.

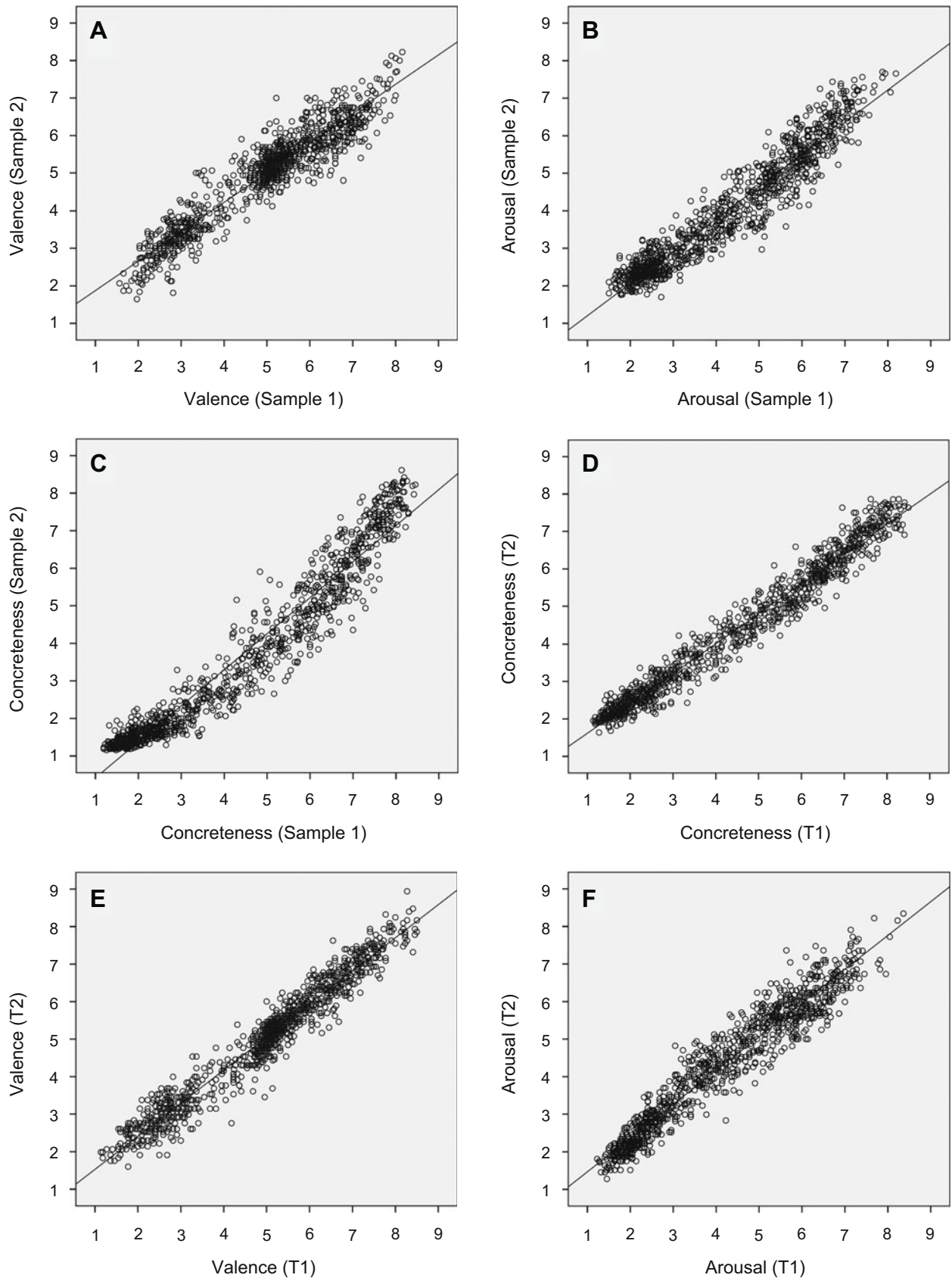


Figure 2. Mean valence (A, E), arousal (B, F), and concreteness (C, D) ratings for each word by Samples 1 and 2 (A, B, C) and by Sample 2 at T1 and T2.

**Table 2**  
**Reliability Measures: Correlations Between the Ratings**  
**of Samples 1 and 2 for All Participants and for Females**  
**and Males Separately, As Well As Correlations**  
**of the Ratings of Sample 1 at T1 and T2**

	Sample 1–Sample 2			Sample 1
	All	Female	Male	T1–T2
Valence	.94**	.87**	.89**	.97**
Arousal	.96**	.91**	.93**	.96**
Concreteness	.97**	.95**	.95**	.98**

\*\* $p < .001$ .

## DISCUSSION

The present study aimed to probe the reliability of normative data regarding valence, arousal, and concreteness ratings in word stimuli. Test–retest reliability, as well as correlations between the ratings of the two samples, were very high, showing that affective and concreteness norms for words can be measured reliably. These data justify the use of word stimuli selected on the basis of affective norms for the experimental study of emotion.

Valence and arousal ratings displayed the typical quadratic function that has also been observed in other normative studies with words and pictorial stimuli (Eilola & Havelka, 2010), indicating that highly negative and positive words were more likely to be rated as highly arousing, whereas neutral words were more likely to be rated as low in arousal. This demonstrates the comparability of the present results to previously established norms. We used this dimensional approach because it seems well suited for the characterization of word stimuli, for which it is difficult to always assign a primary emotional category (e.g., to words such as *bomb*, *vacation*, and *pizza*). However, we acknowledge attempts to use discrete emotional categories (Stevenson et al., 2007), and future studies could validate affective norms in dimensional and categorical ratings.

In comparison with other databases on German words (Lahl et al., 2009; Vö et al., 2009; Vö et al., 2006), the main advantage of the present norms is their high reliability across different samples and across different time points. Test–retest reliability has not been acquired for any other normative word data. It remains an open question, however, whether the high reliability of the present norms is specific to the present database or indicates that word norms are generally highly reliable. One could argue for the latter, because the instructions and methodology used in the different ratings studies are very similar. Nevertheless, it is an empirical question and should be addressed in future research, because it would be very interesting to compare test–retest reliability of different databases.

To conclude, the present study provides a comprehensive database of German nouns with a wide range in emotional status and concreteness. Subsamples of these words have already been successfully used to study the neural basis of emotional processing with functional magnetic resonance imaging and electroencephalography (Kanske & Kotz, in press). Beyond the control of lexical characteristics, the database also allows their systematic

manipulation to investigate, for example, the interaction of emotion and frequency of usage (Nakic et al., 2006), or of emotion and concreteness (Kanske & Kotz, 2007). Furthermore, the stimuli may also be used in clinical studies, because previous data indicate altered processing of emotional words in psychopathology (Besnier et al., 2009; Herrington et al., 2010; Jänsch, Harmer, & Cooper, 2009). Therefore, we hope the Leipzig Affective Norms for German (LANG) will help researchers in the selection of highly controlled word samples and thereby fuel further experimental studies on emotion.

## AUTHOR NOTE

This work was supported by the German Research Foundation (DFG) through the Graduate Program (1182): Function of Attention in Cognition, University of Leipzig, Germany, and by Grant DFG-FOR-499 to S.A.K., Max Planck Institute for Human Cognitive and Brain Sciences. Both authors are members of the Neurocognition of Rhythm in Communication Group at the Max Planck Institute for Human Cognitive and Brain Sciences. Correspondence concerning this article should be addressed to P. Kanske, Department of Cognitive and Clinical Neuroscience, Central Institute of Mental Health, Square J5, 68159 Mannheim, Germany (e-mail: philipp.kanske@zi-mannheim.de).

## REFERENCES

- ALTARRIBA, J. B., BAUER, L. M., & BENVENUTO, C. (1999). Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words. *Behavior Research Methods, Instruments, & Computers*, **31**, 578–602.
- BESNIER, N., RICHARD, F., ZENDJIDJIAN, X., KALADJIAN, A., MAZZOLA-POMIETTO, P., ADIDA, M., & AZORIN, J. M. (2009). Stroop and emotional Stroop interference in unaffected relatives of patients with schizophrenic and bipolar disorders: Distinct markers of vulnerability? *World Journal of Biological Psychiatry*, **10**, 809–818.
- BRADLEY, M. M., & LANG, P. J. (1994). Measuring emotion: The Self-Assessment Manikin and the semantic differential. *Journal of Behavior Therapy & Experimental Psychiatry*, **25**, 49–59.
- BRADLEY, M. M., & LANG, P. J. (1999). *Affective Norms for English Words (ANEW): Stimuli, instruction manual, and affective ratings* (Tech. Rep. C-1). Gainesville: University of Florida, Center for Research in Psychophysiology.
- DILLON, D. G., RITCHEY, M., JOHNSON, B. D., & LABAR, K. S. (2007). Dissociable effects of conscious emotion regulation strategies on explicit and implicit memory. *Emotion*, **7**, 354–365.
- DUDEN (2000). *Die deutsche Rechtschreibung*. Mannheim: Author.
- EILOLA, T. M., & HAVELKA, J. (2010). Affective norms for 210 British English and Finnish nouns. *Behavior Research Methods*, **42**, 134–140.
- EVIATAR, Z., & ZAIDEL, E. (1991). The effects of word length and emotionality on hemispheric contribution to lexical decision. *Neuropsychologia*, **29**, 415–428.
- HAGER, W., & HASSELHORN, M. (1994). *Handbuch deutschsprachiger Wortnormen*. Göttingen: Hogrefe.
- HERRINGTON, J. D., HELLER, W., MOHANTY, A., ENGELS, A. S., BANICH, M. T., WEBB, A. G., ET AL. (2010). Localization of asymmetric brain function in emotion and depression. *Psychophysiology*, **47**, 442–454.
- HODES, R. L., COOK, E. W., & LANG, P. J. (1985). Individual differences in autonomic response: Conditioned association or conditioned fear? *Psychophysiology*, **22**, 545–560.
- JÄNSCH, C., HARMER, C., & COOPER, M. J. (2009). Emotional processing in women with anorexia nervosa and in healthy volunteers. *Eating Behaviors*, **10**, 184–191.
- KANSKE, P., & KOTZ, S. A. (2007). Concreteness in emotional words: ERP evidence from a hemifield study. *Brain Research*, **1148**, 138–148.
- KANSKE, P., & KOTZ, S. A. (in press). Emotion triggers executive attention: Anterior cingulate cortex and amygdala responses to emotional words in a conflict task. *Human Brain Mapping*. doi:10.1002/hbm.21012.
- KUCHINKE, L., JACOBS, A. M., VÖ, M. L., CONRAD, M., GRUBICH, C., & HERRMANN, M. (2006). Modulation of prefrontal cortex activa-



- tion by emotional words in recognition memory. *NeuroReport*, **17**, 1037-1041.
- LAHL, O., GÖRITZ, A. S., PIETROWSKY, R., & ROSENBERG, J. (2009). Using the World-Wide Web to obtain large-scale word norms: 190,212 ratings on a set of 2,654 German nouns. *Behavior Research Methods*, **41**, 13-19.
- MADDOCK, R. J., GARRETT, A. S., & BUONOCORE, M. H. (2003). Posterior cingulate cortex activation by emotional words: fMRI evidence from a valence decision task. *Human Brain Mapping*, **18**, 30-41.
- MATHEWSON, K. J., ARNELL, K. M., & MANSFIELD, C. A. (2008). Capturing and holding attention: The impact of emotional words in rapid serial visual presentation. *Memory & Cognition*, **36**, 182-200.
- NAKIC, M., SMITH, B. W., BUSIS, S., VYTHILINGAM, M., & BLAIR, J. R. (2006). The impact of affect and frequency on lexical decision: The role of the amygdala and inferior frontal cortex. *NeuroImage*, **31**, 1752-1761.
- OLDFIELD, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, **9**, 97-98.
- ORTIGUE, S., MICHEL, C. M., MURRAY, M. M., MOHR, C., CARBONNEL, S., & LANDIS, T. (2004). Electrical neuroimaging reveals early generator modulation to emotional words. *NeuroImage*, **21**, 1242-1251.
- OSAKA, N., OSAKA, M., MORISHITA, M., KONDO, H., & FUKUYAMA, H. (2004). A word expressing affective pain activates the anterior cingulate cortex in the human brain: An fMRI study. *Behavioural Brain Research*, **153**, 123-127.
- OSTERHOUT, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain & Language*, **59**, 494-522.
- PERANI, D., CAPPA, S. F., SCHNUR, T., TETTAMANTI, M., COLLINA, S., ROSA, M. M., & FAZIO, F. (1999). The neural correlates of verb and noun processing: A PET study. *Brain*, **122**, 2337-2344.
- REDONDO, J., FRAGA, I., PADRÓN, I., & COMESAÑA, M. (2007). The Spanish adaptation of ANEW (Affective Norms for English Words). *Behavior Research Methods*, **39**, 600-605.
- RODWAY, P., WRIGHT, L., & HARDIE, S. (2003). The valence-specific laterality effect in free viewing conditions: The influence of sex, handedness, and response bias. *Brain & Cognition*, **53**, 452-463.
- SCOTT, G. G., O'DONNELL, P. J., LEUTHOLD, H., & SERENO, S. C. (2009). Early emotion word processing: Evidence from event-related potentials. *Biological Psychology*, **80**, 95-104.
- SIM, T.-C., & MARTINEZ, C. (2005). Emotion words are remembered better in the left ear. *Laterality*, **10**, 149-159.
- STEVENSON, R. A., MIKELS, J. A., & JAMES, T. W. (2007). Characterization of the Affective Norms for English Words by discrete emotional categories. *Behavior Research Methods*, **39**, 1020-1024.
- STORMARK, K. M., NORDBY, H., & HUGDAHL, K. (1995). Attentional shifts to emotionally charged cues: Behavioral and ERP data. *Cognition & Emotion*, **9**, 507-523.
- VAN HOOFF, J. C., DIETZ, K. C., SHARMA, D., & BOWMAN, H. (2008). Neural correlates of intrusion of emotion words in a modified Stroop task. *International Journal of Psychophysiology*, **67**, 23-34.
- VÖ, M. L., CONRAD, M., KUCHINKE, L., ÜRTON, K., HOFMANN, M. J., & JACOBS, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods*, **41**, 534-538.
- VÖ, M. L., JACOBS, A. M., & CONRAD, M. (2006). Cross-validating the Berlin Affective Word List. *Behavior Research Methods*, **38**, 606-609.
- WENTURA, D. (2000). Dissociative affective and associative priming effects in the lexical decision task: Yes versus no responses to word targets reveal evaluative judgment tendencies. *Experimental Psychology*, **26**, 456-469.

#### SUPPLEMENTAL MATERIALS

The LANG database and the exact instructions given to the raters may be downloaded from <http://brm.psychonomic-journals.org/content/supplemental>.

(Manuscript received April 8, 2010;  
revision accepted for publication May 5, 2010.)