

Cross-validating the Berlin Affective Word List

MELISSA L.-H. VÖ, ARTHUR M. JACOBS, and MARKUS CONRAD
Freie Universität Berlin, Berlin, Germany

We introduce the Berlin Affective Word List (BAWL) in order to provide researchers with a German database containing both emotional valence and imageability ratings for more than 2,200 German words. The BAWL was cross-validated using a forced choice valence decision task in which two distinct valence categories (negative or positive) had to be assigned to a highly controlled selection of 360 words according to varying emotional content (negative, neutral, or positive). The reaction time (RT) results corroborated the valence categories: Words that had been rated as “neutral” in the norms yielded maximum RTs. The BAWL is intended to help researchers create stimulus materials for a wide range of experiments dealing with the emotional processing of words.

In recent years, more and more studies have used words as stimulus materials for studying the influence of emotional content on implicit and explicit memory tasks (e.g., Danion, Kauffmann-Muller, Grangé, Zimmermann, & Greth, 1995; Maratos, Allan, & Rugg, 2000; M. P. Richardson, Strange, & Dolan, 2004; Siegle, Ingram, & Matt, 2002; Windmann & Kutas, 2001). One advantage of verbal stimuli is the ability to control a series of quantifiable factors known to affect word processing, which is not the case with other stimulus materials, such as pictures or films. However, creating highly controlled word-based materials intended for studying the processing of emotional information requires not only the careful matching of various factors known to influence word perception (see Graf, Nagler, & Jacobs, 2005) but also a reliable measure of emotional content.

The present study was motivated by the diversity of findings regarding the influence of emotional valence on both implicit memory tasks—such as word stem completion (see, e.g., Danion et al., 1995) or lexical decision (e.g., Windmann, Daum, & Güntürkün, 2002)—as well as by controversial findings in explicit memory tasks—such as free recall or recognition (e.g., Danion et al., 1995; M. P. Richardson et al., 2004). The lack of consistent results across studies appears to be due not only to differing memory tasks and dependent variables, but also to the great variability of stimulus materials and the different degrees of stimulus control. For example, a number of studies investigating the effect of emotional valence on word memory did not control for the imageability of words, even though there is evidence that easily imageable words are processed more efficiently and are memorized better

than words that are more difficult to image (e.g., Nittono, Suehiro, & Hori, 2002; Paivio, 1971; J. T. E. Richardson, 1975; Strain & Herdman, 1999). Furthermore, Hager and Hasselhorn (1994) found evidence that emotional variables only slightly correlate with the imageability and concreteness of a word.

For pictures, a variety of databases exist, which contain information such as image agreement, age of acquisition, or concreteness (e.g., Bonin, Boyer, Méot, Fayol, & Droit, 2004; Bonin, Peereman, Malardier, Méot, & Chalard, 2003). Lang, Bradley, and Cuthbert (1995) created the International Affective Picture System (IAPS), which has been widely used as a source of standardized stimulus materials in various experiments concerning the effects of emotional pictures. Similarly, considerable effort has been made to collect important information for the creation of various word databases, such as ratings on variables such as imageability, concreteness, or word associations (e.g., Altarriba, Bauer, & Benvenuto, 1999; Bird, Franklin, & Howard, 2001; Bonin, Méot, et al., 2003; Chiarello, Shears, & Lund, 1999; Cortese & Fugett, 2004; Ferrand, 2001; Hager & Hasselhorn, 1994; New, Pallier, Brysbaert, & Ferrand, 2004; Paivio, Yuille, & Madigan, 1968; Ziegler, Stone, & Jacobs, 1997). Siegle's (1994) Balanced Affective Word List project is one of few databases that contain information on the emotional valence of words, but it lacks imageability ratings.

We report a project to collect emotional valence and imageability ratings for more than 2,200 German words, forming the Berlin Affective Word List (BAWL). The primary purpose of this list is to provide researchers, particularly those conducting experiments in German, with information on the emotional valence and imageability of stimulus materials, so that these factors can be either manipulated or controlled. Furthermore, we cross-validated a selection of these words in a forced choice valence decision task by recording reaction times (RTs) and emotional valence ratings. In the “affective valence identification task” reported by Siegle et al. (2002), not the lexicality but the emotionality of a presented word had to be assigned

M.L.-H.V. is now at Ludwig-Maximilians-Universität München. We thank the many students who participated in our study, especially Annet Badel, who carried out the valence decision task, as well as anonymous reviewers for their helpful suggestions. Correspondence relating to this article may be sent to M. L.-H. Vö, Department of Psychology, Ludwig-Maximilians-Universität München, Leopoldstrasse 13, 80802 Munich, Germany (e-mail: lehoa@psy.uni-muenchen.de).

via three keys, labeled “negative,” “neutral,” or “positive.” However, participants in our valence decision task were not asked to indicate the emotional valence of a presented word by choosing one of three categories, but rather had to choose from only two reaction possibilities, either “negative” or “positive,” with no allowance made for a “neutral” response. Apart from presumably allowing more automatic processing of affective valence, this binary decision task allowed us to test the hypothesis that words rated “neutral” in our list should yield maximum RTs. At first glance, this might appear to be a trivial finding, because neutral words had no neutral category in which they could be assigned. However, prolonged RTs for neutral words would support the notion of distinct valence groups that can influence overt behavior.

The selection of words was characterized by a high degree of control for factors known to systematically influence word processing, allowing us to examine whether a word’s emotional content itself can influence overt behavior, as assessed by RTs.

METHOD

Participants

A total of 48 students at Katholische Universität Eichstätt-Ingolstadt and Freie Universität Berlin rated the emotional valence of over 2,200 words, and another 40 students at these institutions rated the imageability of the same words. These ratings are the core of the BAWL.

Furthermore, 8 male and 13 female students at Freie Universität Berlin ranging in age from 21 to 26 ($M = 21.45$ years, $SD = 2.13$) participated in the valence decision task. All reported normal or corrected-to-normal vision and were native German speakers. Two participants had to be excluded because of prolonged RTs.

Stimulus Materials

The Berlin Affective Word List. More than 2,200 verbs and nouns were taken from the CELEX database (Baayen, Piepenbrock, & van Rijn, 1993), representing negative, neutral, and positive affective valences. These words were then rated on both their emotional valence and their imageability by different groups of students, in order to exclude the possibility of mutual influence of the emotional valence and imageability ratings. Emotional valence was rated on a 7-point scale ranging from -3 (*very negative*) through 0 (*neutral*) to $+3$ (*very positive*). The 7-point imageability scale ranged from 1 (*low imageability*) to 7 (*high imageability*). The items were randomly presented to the students to exclude the influence of primacy or recency effects on the participants’ ratings. Subsequently, mean ratings and standard deviations were calculated for each word of the BAWL. Furthermore, the list is characterized by the following variables from the CELEX database (Baayen et al., 1993): Word length ranges from 3 to 10 letters ($M = 6.42$, $SD = 1.58$) and one to four syllables ($M = 2.18$, $SD = 0.67$), the number of phonemes varies from 2 to 10 ($M = 5.58$, $SD = 1.50$), and the total frequency of appearance per million words ranges from low- to high-frequency words ($M = 62.99$, $SD = 164.26$). The database also contains information on the number and frequency of orthographic neighbors ($M = 1.68$, $SD = 2.27$, and $M = 191.92$, $SD = 1,205.30$, respectively), as well as on the number and frequency of higher frequency orthographic neighbors ($M = 0.50$, $SD = 1.08$, and $M = 175.78$, $SD = 1,199.86$, respectively). The complete BAWL can be obtained upon request.

Selection of materials for experimental use. Words were chosen for the experiment as follows.

Targets. From the BAWL, a subset of 180 words was selected according to the following criteria: (1) a mean rating of emotional va-

lence in one of three emotional valence categories—negative (mean emotional valence rating < -1.3), neutral ($-0.8 < \text{mean rating} < 0.8$), or positive (mean rating > 1.3); (2) small standard deviation of the ratings; (3) same number of verbs and nouns within each valence category; and (4) ambiguous words excluded or controlled for. After careful matching, 60 negative, 60 neutral, and 60 positive verbs and nouns were chosen as experimental items. Across the three emotional valence categories, these items did not differ significantly in mean frequency ($M = 32.96$), number of letters ($M = 6.88$), number of syllables ($M = 2.19$), number of orthographic neighbors ($M = 1.33$), number of higher frequency orthographic neighbors ($M = 0.38$), and mean imageability rating ($M = 4.35$). All of these variables were also controlled for across nouns and verbs. In addition, effort was taken to make negative words as “negative” as positive words were “positive”; that is, the mean ratings of negative and positive words were equidistant from the standard neutral value 0 (-1.84 and $+1.84$, respectively), and neutral words had a mean valence rating of 0.02 .

Distractors. The distractors were chosen from the BAWL in the same way as the targets. Again, neither mean frequency ($M = 33.21$), number of letters ($M = 6.98$), number of syllables ($M = 2.30$), number of orthographic neighbors ($M = 1.33$), number of higher frequency orthographic neighbors ($M = 0.33$), nor mean imageability rating ($M = 4.36$) differed significantly between negative, neutral, and positive distractor words. Furthermore, the means of all of these variables did not differ between targets and distractors.

Apparatus

The speeded valence decision task was carried out on an Apple Power Macintosh G4 desktop computer running Mac OS 9.2.2. All stimuli were presented in Courier 24-point type on a 17-in. screen (resolution $1,024 \times 768$ pixels, 85 Hz). The stimuli subtended a vertical visual angle of 0.92° on the screen. The horizontal visual angle ranged from 1.72° for the shortest to 5.72° for the longest word. Stimulus presentation and response recording were controlled by PsyScope experimental software (Cohen & MacWhinney, 1994).

Procedure for the Valence Decision Task

Each participant was seated at a distance of approximately 50 cm from the screen and asked to carefully read the instructions. After a short training period of 6 trials, the 360 experimental trials began. Each word—taken from the BAWL and categorized either as negative, neutral, or positive—was preceded by a fixation cross “+” (300 msec) and stayed centered on the screen until the participant responded. Participants were instructed to assign each word to two distinct emotional valence categories by quickly pressing the “D” or “K” keys, which were labeled “-” (negative) or “+” (positive) for the experiment. The option of pressing “neutral” was not given. RTs were measured from the onset of stimulus presentation until the pressing of one of the two keys. In order to exclude the possibility that RTs could be biased because of the positions of the labels, labels were counterbalanced across participants. The valence decision task lasted about 10 min. The 360 words were then presented to the participants again with the instruction to rate the words’ emotional valence on a scale from -3 to $+3$. Participants completed the rating task in approximately 20 min.

RESULTS

Reaction Time Data

RTs were analyzed in a one-way repeated measures ANOVA, with three levels of emotional valence (negative, neutral, and positive). Participants’ decision times in the valence decision task differed significantly across emotional valence categories [$F(2,18) = 39.34$, $MS_e = 5,942.96$, $p < .01$]. Figure 1 shows the mean RTs for neg-

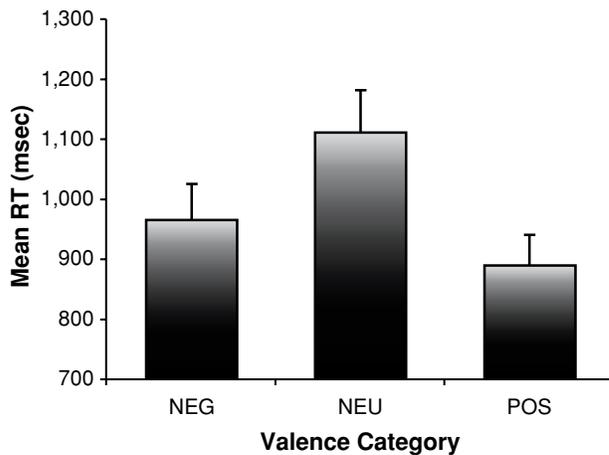


Figure 1. Mean reaction times (RTs, with standard error bars) plotted as a function of valence categories (NEG = negative, NEU = neutral, POS = positive) for the valence decision task.

ative, neutral, and positive words. Planned contrasts revealed significant differences between all three emotional valence categories, with shortest RTs for positive and longest RTs for neutral words [$F(1,18) = 77.32, p < .01$]. RTs for negative words differed significantly from those for neutral words [$F(1,18) = 33.71, p < .01$], as did RTs for positive relative to negative words [$F(1,18) = 8.92, p < .01$]. The differences in RTs across valence categories reflect an effect of emotional valence on the speeded assignment of valence to two discrete categories.

Rating Data

The new ratings of the 360 words in the three valence categories (negative, neutral, and positive) were compared with the assignments already calculated during stimulus acquisition. As can be seen in Figure 2, a correlational analysis for the new and old valence means showed a significant correlation between the old and new ratings ($y = 0.92x + 0.04, R^2 = .91$) reflecting their high reliability.

Reaction Time and Rating Data

Since each participant first took part in the valence decision task and subsequently was asked to rate the same words, we were able to compare a word's mean rating with its mean RT. Figure 3 shows mean RTs plotted as a function of mean ratings. Negative and positive words show shorter RTs than neutral ones. A polynomial fit (quadratic) revealed the following equation: $y = -55.94x^2 - 19.43x + 1,122.7, R^2 = .48$.

DISCUSSION

In this study, we introduced the BAWL, a list containing emotional valence and imageability ratings for over 2,200 German verbs and nouns, in order to respond to the rising need for emotionally valenced word-based stimulus materials. Especially for German words, such a database had been lacking. The information we provide with the BAWL

will enable researchers to either better control stimulus materials or systematically manipulate them according to their emotional content or imageability. Apart from neurocognitive research, clinical research can also benefit from a German database containing emotional valence ratings. For example, studies have clearly indicated an impairment of emotional information processing in depressed or dysphoric patients (e.g., Dietrich et al., 2000; Siegle et al., 2002).

Furthermore, we employed a valence decision task to collect behavioral data regarding the time scale of forced choice valence assignments of 360 words to two discrete valence categories. The selection of words from the BAWL was controlled for a number of factors that are known to systematically influence word processing. The results show that our three theoretically hypothesized emotional valence categories did indeed influence the speed of classification, since RTs differed significantly across the three valence categories. Participants were fastest when categorizing positive words, were slower in categorizing negative words, and showed the longest RTs when neutral words had to be assigned to either a positive or negative valence category. Unlike in Siegle et al. (2002), our participants were not given the possibility of classifying a neutral word as "neutral." The relation between emotional valence ratings and RT latencies was characterized by an inverted, U-shaped function. Furthermore, the emotional valence ratings that followed the valence decision task were highly correlated with those from the BAWL, indicating that the mean ratings for words in the list are valid over time and across participants.

In summary, the information provided by the BAWL is intended to help researchers create stimulus materials for implicit and explicit memory experiments, neuropsychological and clinical studies, and many other applications.

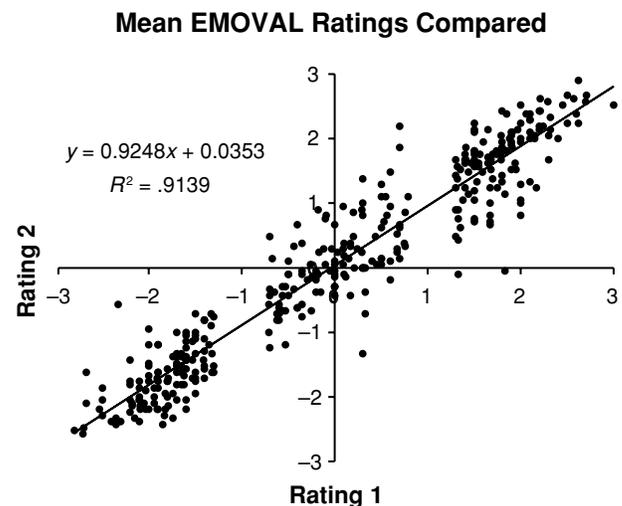


Figure 2. Mean valence ratings of the valence decision task (Rating 2) plotted as a function of mean valence ratings taken from the Berlin Affective Word List (Rating 1).

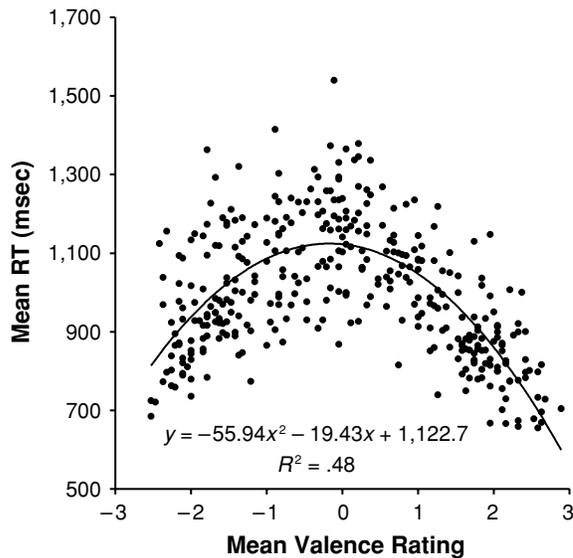


Figure 3. Mean reaction times (RTs) plotted as a function of mean valence ratings for the valence decision task.

REFERENCES

- ALTARRIBA, J., 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.
- BAAYEN, R. H., PIEPENBROCK, R., & VAN RIJN, H. (1993). The CELEX lexical database [CD-ROM]. Philadelphia: Linguistic Data Consortium & University of Pennsylvania.
- BIRD, H., FRANKLIN, S., & HOWARD, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers*, *33*, 73-79.
- BONIN, P., BOYER, B., MÉOT, A., FAYOL, M., & DROIT, S. (2004). Psycholinguistic norms for action photographs in French and their relationships with spoken and written latencies. *Behavior Research Methods, Instruments, & Computers*, *36*, 127-139.
- BONIN, P., MÉOT, A., AUBERT, L., MALARDIER, N., NIEDENTHAL, P., & CAPELLE-TOCZEK, M.-C. (2003). Normes de concrétude, de valeur d'imagerie, de fréquence subjective et de valence émotionnelle pour 866 mots [Concreteness, imageability, subjective frequency, and emotionality ratings for 866 words]. *L'Année Psychologique*, *103*, 655-694.
- BONIN, P., PEEREMAN, R., MALARDIER, N., MÉOT, A., & CHALARD, M. (2003). A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies. *Behavior Research Methods, Instruments, & Computers*, *35*, 158-167.
- CHIARELLO, C., SHEARS, C., & LUND, K. (1999). Imageability and distributional typicality measures of nouns and verbs in contemporary English. *Behavior Research Methods, Instruments, & Computers*, *31*, 603-637.
- COHEN, J. D., & MACWHINNEY, B. (1994). PsyScope [Software]. Pittsburgh: Carnegie Mellon University, Department of Psychology.
- CORTESE, M. J., & FUGETT, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments, & Computers*, *36*, 384-387.
- DANION, J.-M., KAUFFMANN-MULLER, F., GRANGÉ, D., ZIMMERMANN, M.-A., & GRETH, P. (1995). Affective valence of words, explicit and implicit memory in clinical depression. *Journal of Affective Disorders*, *34*, 227-234.
- DIETRICH, D. E., EMRICH, H. M., WALLER, C., WIERINGA, B. M., JOHANNES, S., & MÜNTE, T. F. (2000). Emotion/cognition-coupling in word recognition memory of depressive patients: An event-related potential study. *Psychiatry Research*, *96*, 15-29.
- FERRAND, L. (2001). Normes d'associations verbales pour 260 mots "abstraites" [Word association norms for 260 "abstract" words]. *L'Année Psychologique*, *101*, 683-721.
- GRAF, R., NÄGLER, M., & JACOBS, A. M. (2005). Faktorenanalyse von 57 Variablen der visuellen Worterkennung [Factor analysis of 57 variables in visual word recognition]. *Zeitschrift für Psychologie*, *213*, 205-218.
- HAGER, W., & HASSELHORN, M. (1994). Über Variablen, die eingeschätzt werden sollen, und über Variablen, die eingeschätzt werden: Emotionalität, Angenehmheit, Prägnanz, Erwünschtheit und Sympathie [On variables that should be estimated and variables that are estimated: Emotionality, pleasantness, meaningfulness, desirability, and likability]. In W. Hager & M. Hasselhorn (Eds.), *Handbuch deutschsprachiger Wortnormen* (pp. 226-248). Göttingen: Hogrefe.
- LANG, P. J., BRADLEY, M. M., & CUTHBERT, B. N. (1995). *The International Affective Picture System (IAPS)*. Gainesville, FL: University of Florida.
- MARATOS, E. J., ALLAN, K., & RUGG, M. D. (2000). Recognition memory for emotionally negative and neutral words: An ERP study. *Neuropsychologia*, *38*, 1452-1465.
- NEW, B., PALLIER, C., BRYSSBAERT, M., & FERRAND, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers*, *36*, 516-524.
- NITTONO, H., SUEHIRO, M., & HORI, T. (2002). Word imageability and N400 in an incidental memory paradigm. *International Journal of Psychophysiology*, *44*, 219-229.
- PAIVIO, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart & Winston.
- PAIVIO, A., YUILLE, J. C., & MADIGAN, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monographs*, *76*(1, Pt. 2), 1-25.
- RICHARDSON, J. T. E. (1975). The effect of word imageability in acquired dyslexia. *Neuropsychologia*, *13*, 281-288.
- RICHARDSON, M. P., STRANGE, B. A., & DOLAN, R. J. (2004). Encoding of emotional memories depends on amygdala and hippocampus and their interactions. *Nature Neuroscience*, *7*, 278-285.
- SIEGLE, G. J. (1994). The Balanced Affective Word List Creation Program. Available at www.sci.sdsu.edu/CAL/wordlist.html.
- SIEGLE, G. J., INGRAM, R. E., & MATT, G. E. (2002). Affective interference: An explanation for negative attention biases in dysphoria? *Cognitive Therapy & Research*, *26*, 73-88.
- STRAIN, E., & HERDMAN, C. M. (1999). Imageability effects in word naming: An individual differences analysis. *Canadian Journal of Experimental Psychology*, *53*, 347-359.
- WINDMANN, S., DAUM, I., & GÜNTÜRKÜN, O. (2002). Dissociating pre-lexical and postlexical processing of affective information in the two hemispheres: Effects of the stimulus presentation format. *Brain & Language*, *80*, 269-286.
- WINDMANN, S., & KUTAS, M. (2001). Electrophysiological correlates of emotion-induced recognition bias. *Journal of Cognitive Neuroscience*, *13*, 577-592.
- ZIEGLER, J. C., STONE, G. O., & JACOBS, A. M. (1997). What is the pronunciation for -ough and the spelling for /u/? A database for computing feedforward and feedback consistency in English. *Behavior Research Methods, Instruments, & Computers*, *29*, 600-618.

(Manuscript received June 2, 2005;
revision accepted for publication September 1, 2005.)