*Behavior Research Methods*
*2009, 41 (1), 13-19*
*doi:10.3758/BRM.41.1.13*

# Using the World-Wide Web to obtain large-scale word norms: 190,212 ratings on a set of 2,654 German nouns

**Olaf Lahl**
*Universität Düsseldorf, Düsseldorf, Germany*
*and Universität Erlangen-Nürnberg, Nürnberg, Germany*

**Anja S. Göritz**
*Universität Erlangen-Nürnberg, Nürnberg, Germany*

**Reinhard Pietrowsky**
*Universität Düsseldorf, Düsseldorf, Germany*

**and**

**Jessica Rosenberg**
*Humboldt Universität Berlin, Berlin, Germany*

This article presents a new database of 2,654 German nouns rated by a sample of 3,907 subjects on three psycholinguistic attributes: concreteness, valence, and arousal. As a new means of data collection in the field of psycholinguistic research, all ratings were obtained via the Internet, using a tailored Web application. Analysis of the obtained word norms showed good agreement with two existing norm sets. A cluster analysis revealed a plausible set of four classes of nouns: abstract concepts, aversive events, pleasant activities, and physical objects. In an additional application example, we demonstrate the usefulness of the database for creating parallel word lists whose elements match as closely as possible. The complete database is available for free from ftp://ftp.uni-duesseldorf.de/pub/psycho/lahl/WWN. Moreover, the Web application used for data collection is inherently capable of collecting word norms in any language and is going to be released for public use as well.

Probably due to their omnipresence in everyday life and their direct availability for experimental use, words belong to the most widely used stimuli in cognitive psychology and neighboring disciplines. As a consequence, much effort has been spent by researchers and methodologists in the past to provide word pools with predefined sets of psycholinguistic characteristics known to influence subjects' ability to perceive, process, and retain those words. In this vein, over the last 40 years, measures of concreteness/imageability, frequency/familiarity, valence/emotionality, and age of acquisition have been made available to an international audience, spanning different languages including Chinese (Liu, Shu, & Li, 2007), English (Altarriba, Bauer, & Benvenuto, 1999; G. D. A. Brown, 1984; W. P. Brown & Ure, 1969; Clark & Paivio, 2004; Cortese & Fugett, 2004; Friendly, Franklin, Hoffman, & Rubin, 1982; Gilhooly & Logie, 1980; Kerr & Johnson, 1991; Paivio, Yuille, & Madigan, 1968; Stadthagen-Gonzalez & Davis, 2006; Stevenson, Mikels, & James, 2007; Toglia & Battig, 1978), German (Võ, Jacobs, & Conrad, 2006), Italian (Barca, Burani, & Arduino, 2002), Portuguese (Marques, Fonseca, Morais, & Pinto, 2007), and Spanish (Algarabel, Ruiz, & Sanmartin, 1988; Izura, Hernández-Muñoz, & Ellis, 2005; Pérez & Navalón, 2005).

With the present article, we aim to add to this compilation a set of word norms for 2,654 German nouns. Two objectives informed this norm study: First, to our knowledge, only two data sources for German word norms are currently available, for only a limited number of either stimuli or attributes (Hager & Hasselhorn, 1994; Võ et al., 2006). Second, in the majority of existing norm studies, the rating process is typically performed by relatively few subjects on a large number of stimuli, potentially causing effects of fatigue and stereotyped answer behavior as total testing time increases. Here, by contrast, we made use of the World-Wide Web as a global platform to include a large number of subjects, and each subject needed to spend only a little time and effort. Using this approach, Balota, Pilotti, and Cortese (2001) recently were able to obtain valid estimates on subjective word frequency, where Web-based ratings showed good agreement with traditional paper-and-pencil ratings. The present study is

**O. Lahl, olaf.lahl@uni-duesseldorf.de**

**Table 1**
**Factor Analysis Results for 1,814 Words From Hager and Hasselhorn (1994)**

| Attribute | Factor 1 | Factor 2 | Factor 3 |
|---|---|---|---|
| Imageability | **.94** | .02 | −.04 |
| Concreteness | **.92** | −.05 | −.09 |
| Meaningfulness | **.82** | .10 | .22 |
| Valence | .05 | −.03 | **.95** |
| Arousal | .07 | **.91** | −.25 |
| Potency | −.02 | **.78** | **.50** |

Note—Extraction method: principal component analysis with Kaiser–Guttman criterion. Rotation method: varimax rotation. Absolute factor loadings >.30 are printed in bold.

therefore also intended as a further methodological assessment of using the Internet as a medium to obtain valuable norm data. In what follows, we will describe in detail the collection and analysis of a total of 190,212 ratings on three psycholinguistic attributes from a sample of 3,907 subjects on a word set of 2,654 German nouns, together with an application example for using these norm data.

## METHOD

### Attributes

A factor analysis of the Hager and Hasselhorn (1994) ratings on imageability, concreteness, meaningfulness (cf. Paivio et al., 1968), valence, arousal, and potency (cf. Osgood & Suci, 1955) on a set of 1,814 German nouns, verbs, and adjectives motivated our selection of the psycholinguistic attributes to be rated. According to the three-factor solution shown in Table 1, imageability, concreteness, and meaningfulness all measure a single construct, which we will refer to as *concreteness* for simplicity. Similarly, valence and arousal load on only one of the two other factors each. Potency, however, clearly is an ambiguous variable loading high on both the second and the third factors. For the study at hand, we therefore restricted the set of attributes to the three unambiguous factors of concreteness, valence, and arousal. In addition, word length (number of letters) was directly computed from the words, and word frequency was taken from the CELEX database (Baayen, Piepenbrock, & Gulikers, 1995).

### Word Corpus

A set of 2,654 German nouns was drawn from an electronic corpus of the Institute of German Language and Linguistics of the Humboldt University of Berlin. This corpus contains a collection of several subcorpora that were gathered from daily newspapers, magazines, and common textbooks. Care was taken to exclude verb-derived nouns and obsolete or uncommon nouns, as well as foreign words and words with multiple meanings.

### Procedure

A Web application developed by the first author was used for data collection. On loading the start page of the application, the software selected one out of the three attributes at random and generated a random sample of 50 nouns for the current subject. This way, each subject received a different sample of 50 words that he or she was asked to rate on only one of the three attributes to prevent potentially confusing context switches between the different constructs.

Moreover, the start page (Figure 1A) presented instructions on how to rate the words on the preselected attribute. No mention was made with regard to the two other, nonselected attributes. The subjects were asked to rate the set of 50 words on an 11-point scale ranging from 0 to 10, with higher numbers denoting higher values of the respective variable. To prevent entry of bogus data for word meanings that the subjects felt were unfamiliar or ambiguous, the subjects were also instructed to alternatively click a button labeled "unknown or ambiguous word," which was available for each individual word. The instructions were accompanied by at least two extreme examples to illustrate the concept (e.g. *banana* vs. *being* for the concreteness scale). The subjects were informed that they could access these instructions at any time by clicking a help button, which was always visible.

After indicating their sex, year of birth, native language, and highest educational achievement on the second page (Figure 1B), the subjects rated the 50 words, spanning five pages. Each page (Figure 1C) presented 10 rows with words and rating scales. Advancing to the next page was not possible unless ratings (including the option for unknown or ambiguous words) were made for all the words on a page. At the end, the subjects were thanked for their interest and were dismissed.

### Data Collection

Data were collected from January 14 to May 4, 2008. The subjects were recruited in three ways: (1) e-mail distribution at eight German universities, (2) advertisement on the German Web site www.psychologie-forum.de, which is a noncommercial forum for psychological and related topics, and (3) e-mailing a total of 5,484 members of the online access panel maintained by the second author. The subjects drawn from (1) and (2) took part anonymously, whereas the subjects from (3) were registered users who were identified on log-on by a unique identification number (Panel ID) appended to the URL of the start page.

### Data Integrity

Various checks were performed to ensure the integrity of the entered data. Specifically, a data set was excluded from further analysis if one or more of the following conditions were met. (1) The entire application was completed in less than 90 sec. (2) More than 75% of the ratings had one and the same value. (3) A subject took part multiple times. In this case, only the data from the first participation were used. Recurring participation could explicitly be detected for panel members logging on repeatedly with the same identification number. It was assumed for nonpanelists if data sets with concordant demographic details were entered from a recurring Internet Protocol address and that address was not part of a university network possibly allowing public access of multiple users to the same computer. (4) A user declared being younger than 18.

### Software Environment

The front end of the Web application was programmed in C# using Microsoft (MS) .NET Framework 3.5 Active Server Pages (ASP). MS Internet Information Services (IIS) was used to publish the application on the Web. In addition, a self-developed Tracer application (Lahl & Pietrowsky, 2008) was used to constantly monitor each login and all user actions in real time. All data storage was handled by an MS SQL Server 2005 database. All backend and database programming was done using stored procedures written in Transact Structured Query Language (T-SQL) or C# .NET. More specifically, these routines were used to write user data to the database and to read application data (word set, user interface, display language, etc.) from it. This way, it is generally possible to change an application setting without recompilation.

Data cleansing as described above and data export to MS Excel were automatized by two C# .NET stored procedures. Subjects whose data sets did not fulfill the integrity criteria were not removed from the database but were flagged as invalid, so that their data remained available for potential reanalysis. All statistical analyses were performed with MS Excel 2007 and SPSS 16.0.

### Subjects

From a total of 4,208 completed data sets, 3,907 (2,596 from female subjects, 1,311 from male subjects) were found to be valid according to the criteria outlined above. The mean age of this sample was 31.3 years ($SD$ = 10.8). Within the sample, 95.8% of the

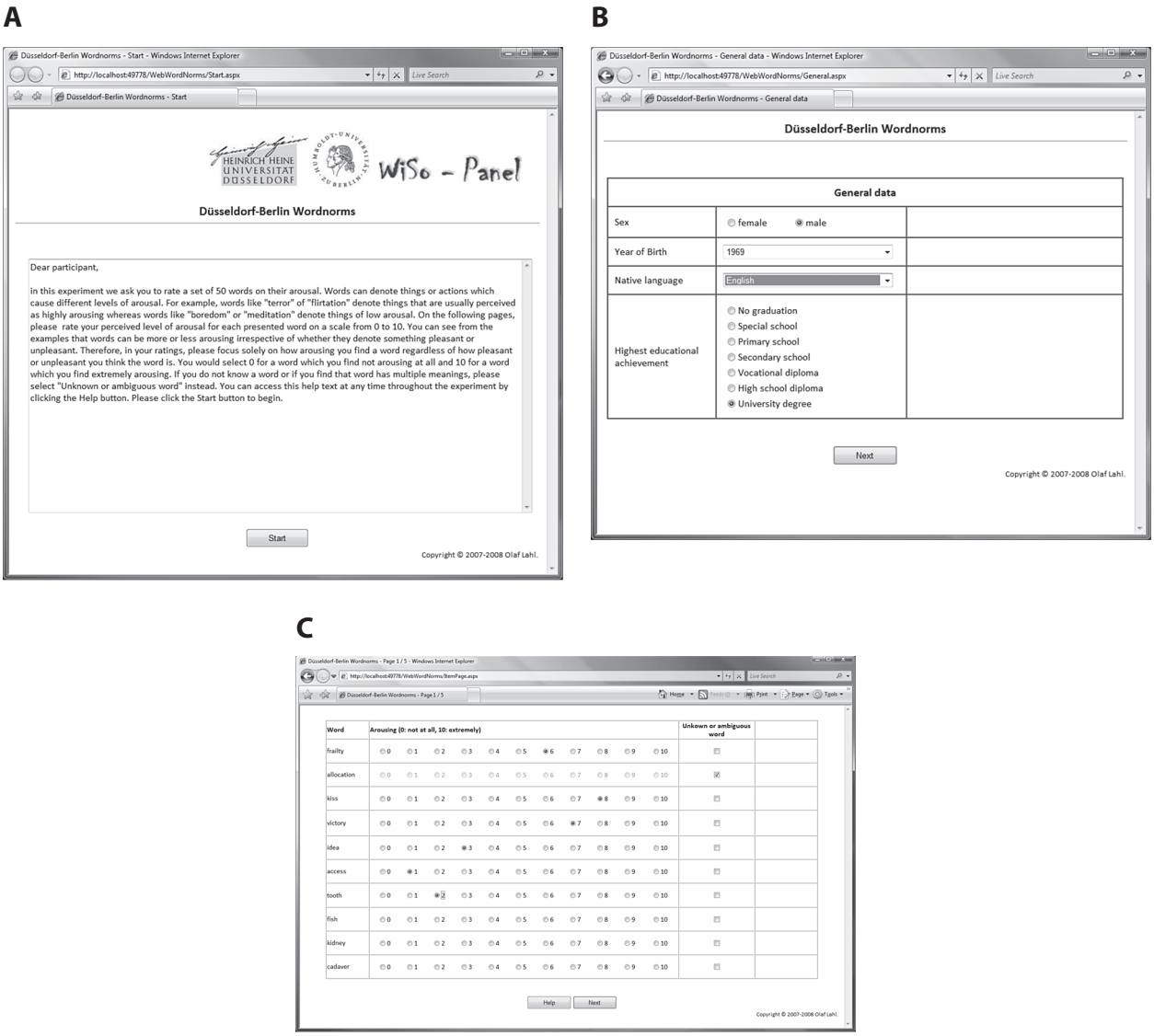


**Figure 1. Screenshots of the Web application used for data collection with welcome page (A), sociodemographic data page (B), and rating page (C). Note that, for the reader's convenience, the English user interface is presented here but that, for the actual data collection, the German user interface was used.**

subjects were native German speakers, and 30.1% held a university degree. From the total sample of 5,484 invited panel members, a subsample of 1,733 users (1,037 of them female, 696 male) accessed the first page of the Web application, yielding a response rate of 31.6% among panel users. Of those, 1,294 produced valid data sets, resulting in a retention rate of 74.7%.

## RESULTS

### Ratings

From Table 2, it can be seen that in the vast majority of cases, the subjects were able to unambiguously identify the words' meanings. There were 190,212 valid ratings on a total of 2,654 words, which means that, on average, 24 ratings are available for each word and attribute. The complete database with sample sizes, mean values, and standard deviations for the three self-rated attributes of each word can be downloaded as an MS Excel file from the first author's ftp site at ftp://ftp.uni-duesseldorf.de/pub/psycho/lahl/WWN.

### Attributes

Figure 2 shows the frequency distributions of the five word characteristics, together with other descriptive statistics. With the three self-rated scales, scores on the left and right tails rarely occur, with the exception of the concreteness scale showing a high frequency of maximum scores. Whereas the distributions for valence and arousal seem to approximate skew-normal distributions, ratings on the concreteness scale appear to be more evenly distributed over the middle and high ranges of the scale. Both the distinct positive skew of the word frequency distribution and the near-normal shape of the word length distribution

**Table 2**
**Numbers of Obtained Ratings**

| Attribute | Number of Ratings | | |
|---|---|---|---|
| | Total | Valid[a] | Valid (%) |
| Concreteness | 60,600 | 58,304 | 96.2 |
| Valence | 70,550 | 68,719 | 97.4 |
| Arousal | 64,200 | 63,189 | 98.4 |
| *Total* | 195,350 | 190,212 | 97.4 |

[a]A rating is defined as valid if it differs from "unknown or ambiguous word."

**Table 3**
**Correlations Between Psycholinguistic Characteristics of 2,654 Nouns**

| Characteristic | Valence | Arousal | Frequency | Length |
|---|---|---|---|---|
| Concreteness | .09* | −.23* | −.03 | −.39* |
| Valence | | −.19* | .13* | −.07* |
| Arousal | | | .02 | .18* |
| Frequency | | | | −.13* |

*$p < .001$.

represent properties of the underlying basic population of nouns.

Table 3 summarizes the linear correlations between the different measures. Coefficients range well below a magnitude of .30, except for concreteness and word length, which show a correlation of −.39. A correlation of similar magnitude between these two variables ($r = -.35$) can be found for the noun data provided by Hager and Hasselhorn (1994).

Assuming that, for a given word and attribute, the intersubject agreement of ratings is an indicator of the general ease and clearness with which subjects are able to make these ratings, we were also interested in systematic changes of measurement heterogeneity as a function of measurement level. In other words, we expected that subjects would more easily (i.e., with lower variance) rate those words with extreme characteristics than those with moderate ones. Figure 3 shows the scatterplots for the standard deviations as a function of the mean values for the three self-rated attributes, together with test statistics for linear and quadratic data fits. Apart from effects of regression to the mean and apart from the fact that, due to the large size of the word pool, all correlations significantly deviate from zero, different patterns for the valence attribute, on the one hand, and for the concreteness and arousal attributes, on the other, may easily be observed. Whereas the standard deviations of the valence ratings only marginally depend on their mean values, there are stronger linear and curvilinear relationships between these statistics for the arousal and concreteness ratings. In particular, subjects show best agreement for words of low arousal and high concreteness.

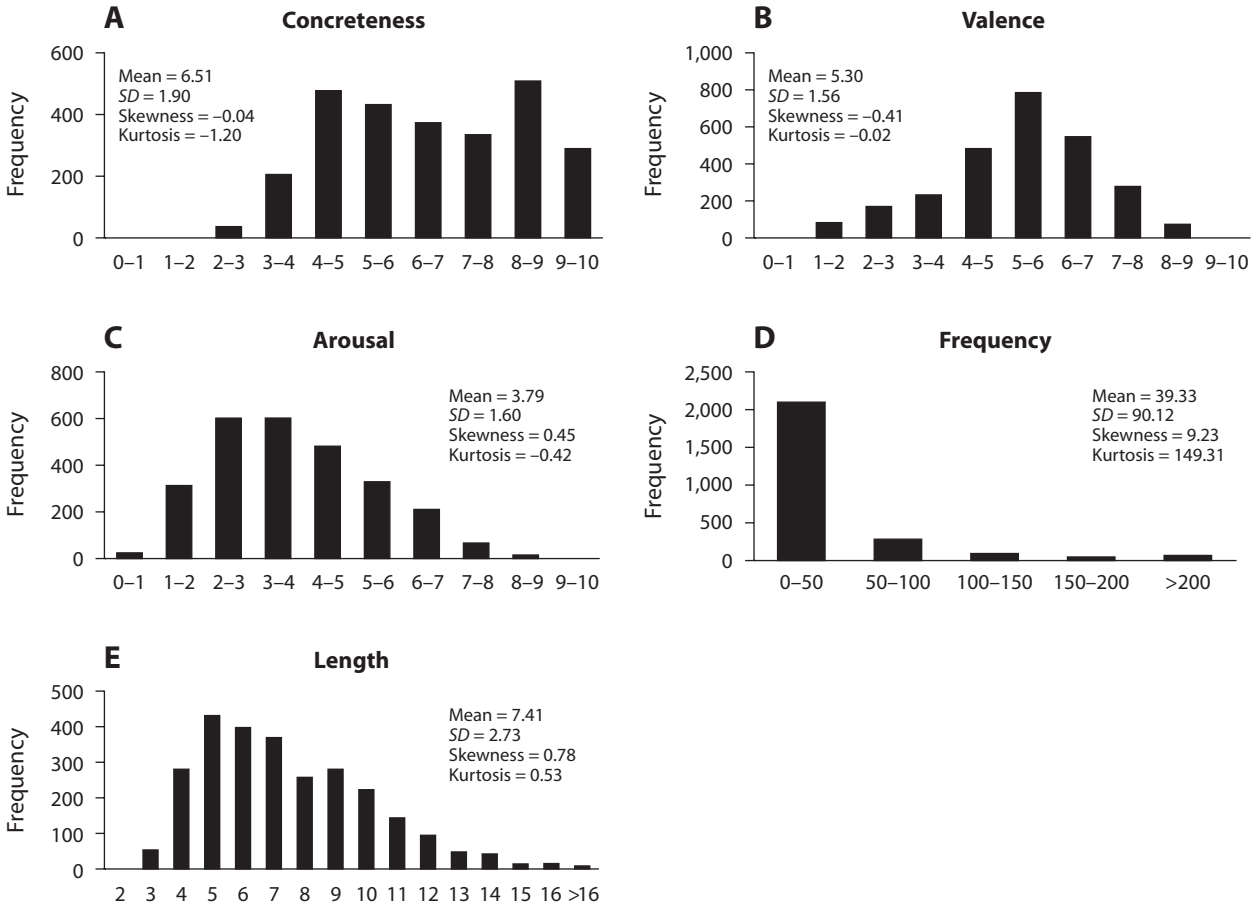


**Figure 2. Frequency distributions of concreteness, valence, arousal, word frequency, and word length.**

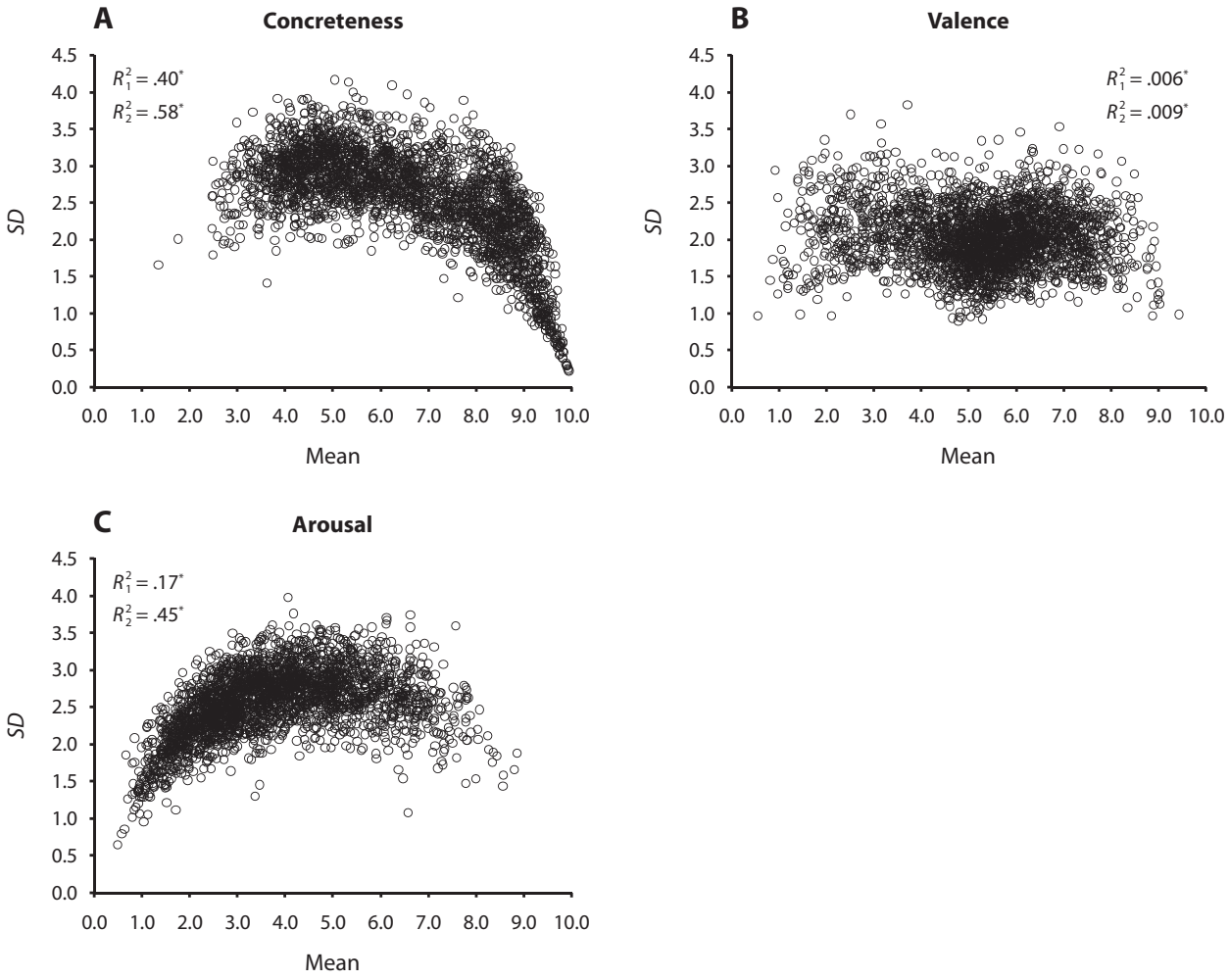


**Figure 3. Plots and test statistics for first- and second-order polynomial fits of the standard deviation of ratings from different subjects as a function of the mean for the three self-rated psycholinguistic attributes of concreteness, valence, and arousal.** $^*p < .001$.

## Validation

To test the validity of the ratings, correlations with the corresponding attributes from two existing German word norm studies were calculated on the sets of shared words in both sets respectively (Table 4). Whereas these validity coefficients were satisfactory for concreteness and valence, the correlation between arousal ratings derived from the present study and those obtained by Hager and Hasselhorn (1994) turned out to be somewhat lower than expected.

## Classification

To check the data set for a plausible classification into subgroups, we computed a cluster analysis over the three self-rated attributes. Table 5 shows the properties of the cluster solution obtained by a two-step cluster analysis using an agglomerative clustering method followed by a partitioning method. In the first step, the Ward algorithm was used to find an adequate set of starting clusters. A set of four clusters was found to be suitable, for the next fusion stage would have almost doubled the increase in

**Table 4**
**Correlations Between Corresponding Attributes of the Present Norm Study and Two Recent Norm Studies**

| Attribute | *N* | Concreteness | Valence | Arousal |
|---|---|---|---|---|
| Hager & Hasselhorn (1994) | 689 | | | |
| Concreteness | | .90 | | |
| Imageability | | .91 | | |
| Valence | | | .91 | |
| Arousal | | | | .58 |
| Võ, Jacobs, & Conrad (2006) | 905 | | | |
| Imageability | | .84 | | |
| Emotionality | | | .93 | |

**Table 5**
**Sizes and *z*-Standardized Centers of the Four Clusters Obtained by a Two-Step Cluster Analysis on the Set of 2,654 Nouns**

| Cluster | *N* | Concreteness | Valence | Arousal |
|---|---|---|---|---|
| 1 | 746 | −0.93 | −0.08 | −0.43 |
| 2 | 435 | −0.19 | −1.57 | 1.27 |
| 3 | 963 | 0.99 | 0.20 | −0.69 |
| 4 | 510 | −0.34 | 1.06 | 0.86 |

**Table 6**
**Sample Sets of Word Pairs and Triples With Minimum Distances in a Five-Dimensional Space of Measures on Concreteness, Valence, Arousal, Frequency, and Word Length**

| Word 1 | Word 2 | Word 3 | |
|---|---|---|---|
| | | | Mahalanobis distance |
| Gepäck (baggage) | Hammer (hammer) | | 0.033 |
| Knospe (bud) | Schaum (foam) | | 0.065 |
| Besserung (amelioration) | Spielraum (elbow room) | | 0.073 |
| Huhn (chicken) | Zaun (fence) | | 0.075 |
| Passage (passage) | Strophe (strophe) | | 0.081 |
| | | | Variance |
| Anker (anchor) | Birne (pear) | Wiege (cradle) | 0.010 |
| Huhn (chicken) | Zaun (fence) | Herd (cooker) | 0.010 |
| Ampel (stoplight) | Raupe (caterpillar) | Zwerg (dwarf) | 0.014 |
| Creme (cream) | Eiche (oak) | Teich (pond) | 0.018 |
| Direktion (directorate) | Intervall (interval) | Kehrseite (downside) | 0.020 |

Note—English translations in parentheses.

the sum of squared errors. This four-cluster solution was then optimized by a subsequent *k*-means cluster analysis. The individual cluster membership of each word can be obtained from the complete database.

As can be seen from Table 5, the first cluster contains nouns of low concreteness, neutral valence, and low arousal. These words usually represent abstract processes and operations (*exchange*, *trade*) or general concepts (*effect*, *class*).[1] The second cluster clearly is a cluster of negative emotions containing nouns that refer mainly to highly aversive events (*loss*, *horror*, *murder*). As a consequence, some swear words (*bitch*) can also be found in this cluster. Cluster 3 (the largest cluster) resembles Cluster 1 with respect to valence and arousal but differs markedly in that it contains mainly words of high concreteness. These words typically refer to physical objects of everyday life (*traffic light*, *curtain*). Finally, the nouns in Cluster 4 denote rather general concepts and activities sharing a positive connotation (*democracy*, *variety*, *hobby*).

### Application Example

Within-subjects designs, as well as parallel test forms, require different sets of stimulus materials that are *matched* on all relevant variables. With regard to word lists, this points to the necessity of creating multiple lists that differ in their semantics but match as closely as possible in their psycholinguistic attributes. To create such equivalent word lists, researchers must be equipped not only with a comprehensive database of word norms, but also with a software application that implements the available matching algorithms. Table 6 contains samples from a set of matched word pairs and triples that we created by applying a software solution provided by Lahl and Pietrowsky (2006) to the norm database presented here. As a supplement to the database, the complete list of tuples may be downloaded from the first author's ftp site. It is intended as a useful application example and to provide researchers with a set of equalized words suitable for use in repeated measures designs. For a recent empirical study in which such equated word lists were successfully adopted, the reader may refer to Lahl, Wispel, Willigens, and Pietrowsky (2008).

## DISCUSSION

This article has introduced a new set of word norms for German nouns that was entirely collected via the Internet. During the last decade, Internet-based data collection has become a popular research method within the behavioral sciences, whose distinct assets and drawbacks have been discussed extensively in two recent articles (Birnbaum, 2004; Reips, 2000). One of the most salient advantages of Internet-based surveys—the ability to gather a large amount of sample data—was of crucial importance for the present study because of the large size of the word corpus under investigation. Obviously, this study, with its extensive word corpus and sample size, would not have been feasible by relying on conventional laboratory methods.

A frequently proclaimed concern about Web-collected behavioral data is a presumed lack of data quality or validity. Although data quality is a concept with many facets, criterion-related validity can easily be assessed by correlating the data obtained with the new method (Internet) with those obtained with a standard method (laboratory). Krantz, Ballard, and Scher (1997) were the first to present a study in which data collected in the laboratory and via the Web measured the same thing. Subsequently, a similar case could be made for other studies and for other research topics (see Krantz & Dalal, 2000, for a review), including the field of psycholinguistic research (Balota et al., 2001). Here, related scales for our new word norms and those from two recent word norm studies proved to correlate satisfyingly ($r > .83$), with the exception of the arousal measure in the present study and a study reported by Hager and Hasselhorn (1994; $r = .58$). A reason for this discrepancy might be sought in the different rating instructions given to subjects. In our instructions for rating words on arousal, special care was taken to prevent the subjects from confounding arousal with valence:

> You can see from the examples that words can be more or less arousing irrespective of whether they denote something pleasant or unpleasant. Therefore, in your ratings, please focus solely on how arousing you find a word, regardless of how pleasant or unpleasant you think the word is.

Unfortunately, Hager and Hasselhorn did not report the instructions given to their subjects for carrying out the arousal ratings. We can therefore only speculate that the modest level of agreement is due to the fact that Hager and Hasselhorn did not include such a caveat in their instructions or is due to some other substantial difference in instructing subjects.

The analysis of measurement heterogeneity as a function of measurement level revealed an interesting dissociation between the valence attribute, on the one hand, and the concreteness and arousal attributes, on the other. Apparently, subjects share a common idea of what is highly concrete and low arousing but show more individual differences when it comes to judging things or actions as pleasant or unpleasant.

The word norms and lists of matched word tuples provided here are intended to facilitate research in all areas of the behavioral sciences that rely on verbal material. Although this freely available material is currently confined to German nouns, the software used for data collection is inherently capable of handling any language and is planned to be made publicly available soon. Equipped with a multilingual Web application for norm data collection and the appropriate tools for equalizing words on multiple variables (Lahl & Pietrowsky, 2006; van Casteren & Davis, 2007), researchers from different countries will then be able to collect comprehensive sets of word norms and carefully matched word sets in any language.

## AUTHOR NOTE

We thank Anke Lüdeling for giving us access to the electronic corpora. Correspondence concerning this article or obtaining the database should be addressed to O. Lahl, Institut für Experimentelle Psychologie, Heinrich-Heine-Universität Düsseldorf, Universitätsstrasse 1 (Geb. 23.03), 40225 Düsseldorf, Germany (e-mail: olaf.lahl@uni-duesseldorf.de).

## REFERENCES

Algarabel, S., Ruiz, J. C., & Sanmartin, J. (1988). The University of Valencia's computerized word pool. *Behavior Research Methods, Instruments, & Computers*, **20**, 398-403.

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., & Gulikers, L. (1995). *The CELEX lexical database* (Release 2) [CD-ROM]. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition*, **29**, 639-647.

Barca, L., Burani, C., & Arduino, L. S. (2002). Word naming times and psycholinguistic norms for Italian nouns. *Behavior Research Methods, Instruments, & Computers*, **34**, 424-434.

Birnbaum, M. H. (2004). Human research and data collection via the Internet. *Annual Review of Psychology*, **55**, 803-832.

Brown, G. D. A. (1984). A frequency count of 190,000 words in the London–Lund Corpus of English Conversation. *Behavior Research Methods, Instruments, & Computers*, **16**, 502-532.

Brown, W. P., & Ure, D. M. J. (1969). Five rated characteristics of 650 word association stimuli. *British Journal of Psychology*, **60**, 233-249.

Clark, J. M., & Paivio, A. (2004). Extensions of the Paivio, Yuille, and Madigan (1968) norms. *Behavior Research Methods, Instruments, & Computers*, **36**, 371-383.

Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments, & Computers*, **36**, 384-387.

Friendly, M., Franklin, P. E., Hoffman, D., & Rubin, D. C. (1982). The Toronto Word Pool: Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1,080 words. *Behavior Research Methods & Instrumentation*, **14**, 375-399.

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation*, **12**, 395-427.

Hager, W., & Hasselhorn, M. (1994). *Handbuch deutschsprachiger Wortnormen*. Göttingen: Hogrefe.

Izura, C., Hernández-Muñoz, N., & Ellis, A. W. (2005). Category norms for 500 Spanish words in five semantic categories. *Behavior Research Methods*, **37**, 385-397.

Kerr, N. H., & Johnson, T. H. (1991). Word norms for blind and sighted subjects: Familiarity, concreteness, meaningfulness, imageability, imagery modality, and word associations. *Behavior Research Methods, Instruments, & Computers*, **23**, 461-485.

Krantz, J. H., Ballard, J., & Scher, J. (1997). Comparing the results of laboratory and World-Wide Web samples on the determinants of female attractiveness. *Behavior Research Methods, Instruments, & Computers*, **29**, 264-269.

Krantz, J. H., & Dalal, R. (2000). Validity of Web-based psychological research. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 35-60). San Diego: Academic Press.

Lahl, O., & Pietrowsky, R. (2006). EQUIWORD: A software application for the automatic creation of truly equivalent word lists. *Behavior Research Methods*, **38**, 146-152.

Lahl, O., & Pietrowsky, R. (2008). Tracer: A general purpose software library for logging events in computerized experiments. *Behavior Research Methods*, **40**, 1163-1169.

Lahl, O., Wispel, C., Willigens, B., & Pietrowsky, R. (2008). An ultra short episode of sleep is sufficient to promote declarative memory performance. *Journal of Sleep Research*, **17**, 3-10.

Liu, Y., Shu, H., & Li, P. (2007). Word naming and psycholinguistic norms: Chinese. *Behavior Research Methods*, **39**, 192-198.

Marques, J. F., Fonseca, F. L., Morais, A. S., & Pinto, I. A. (2007). Estimated age of acquisition norms for 834 Portuguese nouns and their relation with other psycholinguistic variables. *Behavior Research Methods*, **39**, 439-444.

Osgood, C. E., & Suci, G. J. (1955). Factor analysis of meaning. *Journal of Experimental Psychology*, **50**, 325-338.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 words. *Journal of Experimental Psychology Monograph Supplement*, **76**(1, Pt. 2).

Pérez, M. A., & Navalón, C. (2005). Objective-AoA norms for 175 names in Spanish: Relationships with other psycholinguistic variables, estimated AoA, and data from other languages. *European Journal of Cognitive Psychology*, **17**, 179-206.

Reips, U. D. (2000). The Web experiment method: Advantages, disadvantages, and solutions. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 89-117). San Diego: Academic Press.

Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods*, **38**, 598-605.

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.

Toglia, M. P., & Battig, W. F. (1978). *Handbook of semantic word norms*. Hillsdale, NJ: Erlbaum.

van Casteren, M., & Davis, M. H. (2007). Match: A program to assist in matching the conditions of factorial experiments. *Behavior Research Methods*, **39**, 973-978.

Võ, M. L.-H., Jacobs, A. M., & Conrad, M. (2006). Cross-validating the Berlin Affective Word List. *Behavior Research Methods*, **38**, 606-609.

## NOTE

1. All samples are best-matching English translations of the corresponding German words.

(Manuscript received June 6, 2008;
revision accepted for publication August 18, 2008.)