Norms of Valence, Arousal, Dominance, and Age of Acquisition for 4300 Dutch Words

Agnes Moors[1], Jan De Houwer[1], Dirk Hermans[2], Sabine Wanmaker[3], Kevin van Schie[3], Anne-Laura Van Harmelen[4], Maarten De Schryver[1], Jeffrey De Winne[1], and Marc Brysbaert[1]

[1] Ghent University

[2] University of Leuven

[3] Erasmus University Rotterdam

[4] Leiden University

Author note: Please address correspondence to Agnes Moors, Department of Psychology, Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium. Email: agnes.moors@ugent.be

Preparation of this paper was supported by Methusalem Grant BOF09/01M00209 of Ghent University to Jan De Houwer.

Abstract

This article presents norms of valence/pleasantness, activity/arousal, power/dominance, and age of acquisition for 4,300 Dutch words, mainly nouns, adjectives, adverbs, and verbs. The norms are based on ratings with a 7-point Likert scale by independent groups of students from two Belgian (Ghent and Leuven) and two Dutch samples (Rotterdam and Leiden-Amsterdam). For each variable, we obtained high split-half reliabilities within each sample and high correlations between samples. In addition, the valence ratings of a previous, more limited study (Hermans & De Houwer, 1994) correlated highly with those of the present study. Therefore, the new norms are a valuable source of information for affective research in the Dutch language.

Keywords: lexical norms, valence, arousal, dominance, age of acquisition, frequency, word length

Norms of Valence, Arousal, Dominance, and Age of Acquisition for 4300 Dutch Words

Psychological research often involves stimuli with an affective meaning, such as words, pictures, odors, noises, stories, and films. We use the term “affective” to refer to an open-ended set of variables that are characteristic of phenomena including emotions, moods, attitudes, evaluation or appraisal, and feelings. These phenomena are related but differ in some respects. Emotions are often thought to consist of changes in multiple components such as evaluation or appraisal, physiological responses, motor expressions (facial, vocal, gestural), action tendencies, and feelings (subjective experience). Moods are considered to have somewhat fewer components or to have less pronounced values for all of these components. Both emotions and moods are characterized by the variables that characterize their components. Examples of variables characterizing the feeling component of emotions and moods are valence, arousal, and power or dominance (Fontaine, Scherer, Roesch, & Ellsworth, 2007). Variables characterizing the appraisal component of emotions and moods are valence, goal relevance, goal congruence, power or coping potential, agency, novelty, and certainty (Ellsworth & Scherer, 2003). The variable most often mentioned as characterizing attitudes is valence.

Research on affective phenomena can be organized into various lines of research. A first line focuses on the processes involved in the production and perception of affective phenomena. This includes research on the processes involved in (a) the formation, activation, and change of attitudes (e.g., Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010), and (b) the components of emotions such as appraisal (e.g., Moors, 2010; Sander, Grandjean, & Scherer, 2005), action tendencies (e.g., Drake & Meyers, 2011), somatic responses (e.g., Bauer, 1998; Lang, Bradley, & Cuthbert, 1998; Phan, Wager, Taylor, & Liberzon, 2002), expressive behavior (e.g., Russell, Bachorowski, & Fernandez-Dols, 2003), and feelings (e.g., Dan Glauser & Scherer, 2008). A second line of research focuses on the relation among

different affective phenomena. This line is concerned with the interrelations among the various emotion components (appraisal, action tendencies, expressive behavior, bodily responses, and feelings; Scherer, 2009; Roseman & Evdokas, 2004). A third line of research focuses on the relation between affective and non-affective phenomena. This includes research on the relations between emotions (or their components), moods, and attitudes, on the one hand, and attention, perception, memory, judgments, and decision making on the other hand (Compton, 2003; Kensinger, 2004; Laney, Campbell, Heuer, & Reisberg, 2004; Levine & Pizarro, 2004; Williams, Mathews & McLeod, 1996; Vohs, Baumeister, & Loewenstein, 2007).

All these lines of research make use of stimuli with an affective meaning. For example, in research on the influence of evaluation on memory, the aim is to manipulate the content of evaluation and to measure its influence on memory. Manipulation of the content of evaluation is accomplished by presenting stimuli that are expected to be evaluated by the participants in a certain way, for example, as positive or negative (e.g., Bower, Gilligan, & Monteiro, 1981). For another example, to examine the influence of evaluations of power on action tendencies, researchers have primed participants with strong and weak words and measured their tendencies to approach and avoid (Smith & Bargh, 2008). Stimulus selection needs to proceed in such a way that researchers can be confident that most participants will evaluate the stimuli in the intended way. The preferred method for stimulus selection is to choose stimuli from previous rating studies.

Such rating studies have been reported for pictures (Lang, Bradley, & Cuthbert, 2008), sounds (Bradley & Lang, 1999b), and words in several languages, including English (Bradley & Lang, 1999a; Brown & Ure, 1969; Stevenson, Mikels, & James, 2007), Spanish (Redondo, Fraga, Padrón, & Comesaña, 2007), French (Bertels, Kolinski, & Morais, 2009; Bonin, Méot, Aubert, Malardier, Niedenthal, & Capelle-Toczek, 2003; Corson & Quistrebert, 2000;

Messina, Moré, & Cantraine, 1989; Niedenthal, Auxiette, Nugier, Dalle, Bonin, & Fayol, 2004; Syssau & Font, 2005), German (Grühn & Smith, 2008; Hager & Hasselhorn, 1994; Lahl, Göritz, Pietrowsky, & Rosenberg, 2009; Võ et al., 2009), and Finish (Eilola & Havelka, 2010). In Dutch, a study by Hermans and De Houwer (1994) provided valence ratings and subjective familiarity ratings for 740 Dutch words of which 370 were adjectives referring to personality traits and 370 were nouns.

The present study provides affective ratings for 4,300 Dutch words. It goes beyond many other word norm studies (in Dutch and other languages) in several respects. We included more words, which belonged to more grammatical categories, and which were tested on more affective variables in more populations. Specifically, the set of 4,300 words consisted mainly of nouns, adjectives, adverbs, and verbs. The words were evaluated on the variables of valence/pleasantness, activity/arousal, and power/dominance. Power/dominance (also sometimes referred to as potency or control) has not often been included in previous word norming studies (but see Bradley & Lang, 1999a), even though it has been identified as an important variable in emotion research in addition to valence and arousal (e.g., Fontaine et al., 2007; Osgood, Suci, & Tannenbaum, 1957). In addition, we collected ratings of age of acquisition (AoA), so that the stimuli can be controlled for or manipulated on this variable as well. We chose AoA because it has been presented as the fifth most important factor determining word recognition times after frequency, word length, similarity to other words, and word onset (Kuperman, Stadthagen-Gonzalez, & Brysbaert, in press). Imageability is another variable that is sometimes controlled or tested, but Brysbaert, Lange, and Van Wijnendaele (2000) found that in Dutch it explains virtually no variance once the words are controlled for frequency and AoA. A list of imageability ratings in Dutch can be found in Van Loon-Vervoorn (1989). It includes ratings for about 67% of the words in the present list.

The ratings were performed by equally sized groups of male and female students from two Belgian (Ghent and Leuven) and two Dutch (Rotterdam and Leiden-Amsterdam) samples, which allowed us to see how region-independent they are. A further strength is that in our study, each participant rated the entire set of words for only one variable. This has the advantage that the ratings for one variable (e.g., valence/pleasantness) could not influence or "contaminate" the ratings for another variable (e.g., activity/arousal or power/dominance; cf. Bestgen & Vincze, 2012; Bradley & Lang, 1994).

Method

Participants

Participants were 224 students (112 women, 112 men) recruited from two universities in Flanders (the Dutch speaking half of Belgium; Ghent University, N = 64; and the University of Leuven, N = 48) and three universities in The Netherlands (Erasmus University Rotterdam, N = 64, Leiden University, N = 41, and the University of Amsterdam, N = 7). The participants of Leiden University and Amsterdam University were treated as one sample. The remaining universities each constituted one sample. Each of the samples consisted of an equal number of women and men. Participants at Ghent and Leuven received 50 euros for their help. In Leiden and Amsterdam, they received 20 euros. In Rotterdam, they received course credits. The age of the participants ranged from 17 to 58 (M = 22.08; SD = 4.49). The ratings were obtained between May 2011 and February 2012. Our participants were students because this is the population typically tested in the studies for which the ratings are meant.

Materials and Procedure

We selected 4,300 Dutch words from various sources (Keuleers, Diependaele, & Brysbaert, 2010; De Deyne & Storms, 2008; Fontaine, Poortinga, Setiadi, & Suprapti, 2002; Fontaine et al., 2007; Frijda, Kuipers, & ter Schure, 1989; Hermans & De Houwer, 1994; Osgood, Suci, & Tannenbaum, 1957; Rouckhout & Schacht, 2000; http://synoniemen.net/).

The selection of words was guided by the idea that in addition to neutral words, we needed as many words as possible with a marked value for each of the three affective variables. The set mostly contained nouns, adjectives, adverbs, and verbs. We excluded most interjections, most plurals, diminutives, words that have become obsolete, words with a very low frequency in written language, and words that were uncommon in either region (Flanders/the Netherlands). Of the 740 words of the rating list of Hermans and De Houwer (1994) 715 were included in the present list. This allowed us to examine whether the valence ratings of these 715 words generalized to the present study.

Each participant rated the entire set of 4,300 words for one variable only: valence/pleasantness, activity/arousal, power/dominance, or AoA. In each sample, each affective variable (valence/pleasantness, activity/arousal, power/dominance) was rated by 8 women and 8 men. AoA was rated only at Ghent University and the Erasmus University Rotterdam (in each university by 8 women and 8 men).To reduce possible sequence effects, the order in which words appeared in the list was randomized for each participant separately.

Participants who accepted to take part in the study, received an email with an Excel file containing two sheets: The first sheet presented the instructions; the second sheet listed the 4,300 words. Samples of these Excel files for each variable are provided as supplementary materials to this article. Participants in the valence/pleasantness condition were asked to judge the extent to which the words in the study referred to something that is positive/pleasant (“positief/aangenaam”) or negative/unpleasant(“negatief/onaangenaam”), using a 7 point-scale (1 = very negative/unpleasant, 2 = fairly negative/unpleasant, 3 = somewhat negative/unpleasant, 4 = neutral, 5 = somewhat positive/pleasant, 6 = fairly positive/pleasant, 7 = very positive/pleasant). To ensure that the participants understood the instructions, we provided the following examples with words that did not appear in the list:

"If you think that "atom bomb" has a very negative meaning, please choose 1. If you think that "fantastic" has a very positive meaning, please choose 7. If you think that "sprouts" refers to something that is fairly unpleasant, please choose 2. If you think that "relaxing" refers to something that is fairly pleasant, please choose 6."

Participants in the activity/arousal condition were asked to judge the extent to which the words in the study referred to something that was active/arousing ("actief/opgewonden") or passive/calm ("passief/kalm"), using a 7 point scale (1 = very passive/calm, 2 = fairly passive/calm, 3 = somewhat passive/calm, 4 = neutral, 5 = somewhat active/aroused, 6 = fairly active/aroused, 7 = very active/aroused). The examples provided for this dimension were:

"If you think that "hammock" has a fairly passive meaning, please choose 2. If you think that "working" has a fairly active meaning, please choose 6. If you think that "meditating" has a very calm meaning, please choose 1. If you think that hyperkinetic has a very aroused meaning, please choose 7."

Participants in the power/dominance condition were asked to judge the extent to which the words in the study referred to something that was weak/submissive ("zwak/onderdanig") or strong/dominant ("sterk/dominant"), using a 7 point scale (1 = very weak/submissive, 2 = fairly weak/submissive, 3 = somewhat weak/submissive, 4 = neutral, 5 = somewhat strong/dominant, 6 = fairly strong/dominant, 7 = very strong/dominant). The examples provided for this variable were:

"If you think that "grass stalk" refers to something that is very weak, please choose 1. If you think that "avalanche" refers to something that is very strong, please choose 7. If you think that "servant" has a fairly submissive meaning, please choose 2. If you think that "revenge" has a fairly dominant meaning, please choose 6."

After reading the instructions, the participant opened the second sheet. The 4,300 words were presented in the first column. The participants rated each word by typing a number from 1 to 7 in the second column. After they had typed a number, the meaning of the number appeared in the third column (e.g., when the participant had pressed 2, the message "fairly passive/calm" appeared). When the participant typed a wrong number (outside of the 1 to 7 range), a red square with the message "wrong code" appeared. Participants were instructed to respond as accurately as possible, but not to think too long. They could type in the letter N when they did not know the word.

The same procedure was used in the AoA condition, except that participants were asked to enter the age at which they thought they had learned the word (Ghyselinck, De Moor, & Brysbaert, 2000; Bird, Franklin, & Howard, 2001). We clarified that this was the age at which they first understood the word when somebody else used it in their presence, even when they did not use the word themselves. The examples given for this variable were:

"If you think you learned "banana" when you were 3 years old, please fill in 3
If you think you learned "accountant" when you were 11 years old, please fill in 11."

The validity of AoA ratings has been confirmed in studies that obtained a high corrrelation between AoA ratings and the percentage of words known by children of various ages (e.g., De Moor, Ghyselinck, & Brysbaert, 2000; Morrison, Chappell, & Ellis, 1997). Participants were asked to send the completed file back via email to the experimenter in approximately 2 weeks. Afterwards, they were invited to collect the monetary reward or course credits.

Results

Outlier Analysis

We conducted the following outlier analysis. First, we discarded all ratings on which participants indicated that the word was unknown to them (1.1%). We then calculated the

mean and SD for each word. Next, we counted for each participant the percentage of words for which their rating deviated 2.5 SDs from the mean. Only one participant (who rated AoA) had a high percentage of outliers (30.8 %) and was discarded. The percentage of outliers for the other participants ranged between 0 and 17.5 (M = 1.4; SD = 2.3). We then calculated the mean and SD for each word a second time on the remaining data. Further, we excluded the ratings for one word because it had been typed incorrectly in the Excel files. Finally, there were 42 missing values on a total of 963,200 ratings. All in all, 947,462 or 98.4% valid ratings were obtained.

Ratings of the Affective Variables and AoA

An Excel file with the raw data is provided as supplementary materials to this article. It contains the 4,300 words in alphabetical order together with their English translation (based on Google Translate and Van Dale Groot Woordenboek) and the mean values (M), standard deviations (SD), and sample sizes (N) for valence/pleasantness (V), activity/arousal (A), power/dominance (P), and age of acquisition (AoA). The file also contains information about word frequency (FR) and number of letters (Let). The frequency scores were taken from the SUBTLEX-NL database (Keuleers, Brysbaert, & New, 2010). The file contains both frequency per million words and log10 of frequency per million words. Forty words in our study did not appear in the SUBTLEX-NL database. Following Brysbaert and New (2009) and Keuleers, Brysbaert, et al. (2010), we assigned values of freq pm = .02 and log10 = -1.64 to these words, in line with the size of the SUBTLEX-NL corpus (43.8 million words).

The data for the first four variables are split into three columns: the data of the global sample (All), followed by those of the women (Women), and those of the men (Men). Further, there is a column with the percentage of participants (across all ratings) who indicated they did not know the word. Finally, we added a column in which each word received a code for the most frequent grammatical category (part of speech) to which it belongs: nouns (N),

adjectives and adverbs (A), verbs (V), and a small rest category with numerals and interjections (R). Researchers are referred to the SUBTLEX-NL file for more information about the words (Keuleers, Brysbaert, et al., 2010; also available online at http://crr.ugent.be/isubtlex/). It may be noted that, in line with most previous research, participants did not receive explicit instructions about ambiguous words. Thus, this ambiguity may be reflected in the rating variability.

Descriptive statistics of the variables are presented in Table 1. Figures 1 to 4 show plots of the means and standard deviations of the ratings (together with the English translations of some outliers) for all dependent variables. The scatterplot for valence/pleasantnesss (Figure 1) shows that there are two types of words in the midrange (around the score of 4): (a) words with low SDs upon which participants agree that they are neutral, and (b) words with high SDs that elicited both high and low values from different participants (examples are “pugnacious” and “complacent”). Inspection of the scatterplot for arousal/activity (Figure 2) shows that there is more consensus about the high arousing and low arousing words than about the words in the midrange (around the score of 4). The scatterplot for power/dominance (Figure 3) is somewhat similar to that of valence, but less pronounced. Finally, the scatterplot of AoA shows that the SDs increase with increasing means. This suggests that participants learn similar words in the first years of life, but show more variability in later years (also recall that participants were not using a Likert scale for this variable).

Reliability

We calculated the split-half reliabilities for each sample separately. Samples were split in halves by using the entrance ranks of the participants (separately for males and females) and making a distinction between the participants with odd and even ranks. For each group we calculated the mean rating for each word, and we then correlated the means of both

groups. As shown in Table 2, the adjusted correlations using the Spearman-Brown formula were very high, ranging from r = .82 to r = .97. Furthermore, we obtained high correlations of at least r =.82 between the samples. The fact that the correlations between samples were as high as the correlations within samples indicates that the ratings of the words were not subject to strong regional differences, meaning that the average values can be used across the entire Dutch-speaking area (remember that we selected words known both in Flanders and the Netherlands).

To further test the generalizability of our ratings, we correlated them with ratings from previous studies. For the valence ratings there were 715 words in common with Hermans and De Houwer (1994). Figure 5 shows a strong linear relationship between the ratings of both studies, r = .96. For AoA, we correlated our ratings with those of Ghyselinck et al. (2000) and Ghyselinck, Custers, and Brysbaert (2003). For the first study, there were 1,307 words in common, and a correlation of r = .93; for the second study, there were 710 words in common and a correlation of r = .95.

Correlations between Variables

Pearson correlations were calculated between the affective variables, AoA, frequency, and word length (i.e., number of letters). No linear relation was found between valence/pleasantness and activity/arousal, but we did obtain a quadratic relation: After centering the mean ratings of valence/pleasantness, we obtained a positive correlation (r = .29) between the square of the centered valence/pleasantness scores and activity/arousal (Figure 6) . Power/dominance had a positive correlation with valence/pleasantness (r= .27; Figure 7) and a high positive correlation with activity/arousal (r= .59; Figure 8). Thus, words rated as more dominant were also rated as more positive and more active.

AoA correlated negatively with valence/pleasantness (r = -.17) and positively with power/dominance (r = .08), suggesting that words that were learned early in life were rated as

more positive and less dominant. No linear relation was found between AoA and activity/arousal (r = .03).

Frequency had a positive correlation with valence/pleasantness (r = .15), activity/arousal (r = .10), and power/dominance (r = .17), indicating that frequent words were rated as more positive, more active, and more dominant.

Word length had a low negative correlation with valence/pleasantness (r = -.08) as well as low positive correlations with activity/arousal (r = .19) and power/dominance (r = .08). This means that longer words were rated as slightly more negative, somewhat more active, and slightly more dominant. AoA had a strong negative correlation with frequency (r = -.60) and a positive correlation with word length (r = .33), indicating that words learned early in life are more frequent and shorter. Frequency and word length (r = -.25) also correlated negatively, which means that more frequent words are shorter. All reported correlations were significant with $p < .001$.

Discussion

We collected word norms for 4,300 Dutch words for the affective variables valence/pleasantness, activity/arousal, and power/dominance, and for AoA. Ratings for the first three variables were performed with 7-point Likert scales; the AoA-ratings reflect the age at which participants think they acquired the words. Ratings were collected at different universities to make sure that they applied to all Dutch-speaking regions. Virtually all words belong to the grammatical categories of nouns, adjectives, adverbs, and verbs. Our study goes beyond previous studies (in Dutch and other languages) in that we obtained ratings on more affective variables, for a larger set of words, covering more grammatical categories, and carried out by more populations.

We observed high split-half reliabilities within samples and equally high correlations between samples, indicating that there is a large agreement among the students within and

between the various samples. We also found that the ratings of previous, more limited studies (Ghyselinck et al., 2000, 2003; Hermans & De Houwer, 1994) generalized to those of the present study. We can therefore conclude that the norms that we obtained are reliable and can be used confidently for the selection of words in affective research.

An exploration of the relations between the affective variables revealed a quadratic relation between valence/pleasantness and activity/arousal. This confirms previous findings of a small but consistent U-shaped relationship between valence and arousal in studies with words (e.g., Bradley & Lang, 1999a; Kanske & Kotz, 2010; Redondo et al., 2007; Võ et al., 2009) and pictures (e.g., Bradley & Lang, 1994; Cuthbert, Bradley, & Lang, 1996).

We also found positive correlations of power/dominance with valence/pleasantness and with activity/arousal. Few previous studies collected ratings for dominance in addition to ratings for valence and arousal, and even fewer studies reported on the relation between dominance and other variables. Studies that did collect ratings for dominance (or related constructs such as potency or control; Bradley & Lang, 1994, 1999a; Grühn & Smith, 2008; see also Keltner, Gruenfeld, & Anderson, 2003) reported positive correlations between dominance and valence. The results are mixed for dominance and arousal: Grühn and Smith (2008; 200 words) reported no correlation (r = -.09, n.s.); an analysis performed by us on the Bradley and Lang (1999a; 1030 words) data revealed a weak positive correlation (r = .07, p = .021); Bradley and Lang (1994; 21 pictures) reported negative correlations (ranging from r = -.14 to r = -.57). Several factors may have contributed to this divergence. First, participants in our study rated the active and dominant meaning of the stimuli whereas participants in the other studies rated their own feelings of activity and dominance in response to the stimuli. Thus, a participant may rate a snake as having an active and dominant meaning but his/her own feelings as active and submissive. This may have played less in the Bradley and Lang (1999a) study because the stimuli were words that referred not only to emotion-eliciting

stimuli (like snakes and injuries) but also to emotional states (like fear and anger). Second, the divergent correlations may be due to differences between the samples of words tested (given that each study presented only a subsample of the words). Bradley and Lang (1994), for instance, collected ratings for only 21 pictures. Like in all rating studies, the correlations obtained reflect the structure of the specific stimulus set used. Larger stimulus sets are more likely to be representative for the universe of stimuli than smaller stimulus sets. Third, participants in our study each rated only one affective variable (i.e., between-subjects design) whereas participants in the other studies rated all affective variables (i.e., within-subjects design). Thus, it could be argued that the participants in our study were focused less on the differences between dominance and arousal than the participants in the other studies. This may explain why we obtained a stronger positive correlation between valence and arousal than the other word rating studies (Bradley & Lang, 1999a; Grühn & Smith, 2008).

Several of the other patterns of correlations that we observed are compatible with previous findings as well. That is, other studies confirmed that words learned early in life are more positive (Citron, Weekes, & Ferstl, 2009), more frequent (Ghyselinck et al., 2000; Morrison et al., 1997; Stadthagen-Gonzales & Davis, 2006; Citron et al., 2009) and shorter (Ferrand et al., 2008), that frequent words are more positive (Grühn & Smith, 2008) and shorter (Ferrrand et al., 2008; Grühn & Smith, 2008), and that high-arousing words are longer (Grühn & Smith, 2008). To conclude, we believe the present study will be a valuable source of information for affective research that makes use of Dutch words.

References

De Deyne S., & Storms G. (2008). Word Associations: Norms for 1,424 Dutch words in a continuous task. *Behavior Research Methods, 40,*198-205.

Bauer, R. M. (1998). Physiologic measures of emotion, 15, 388-396. *Journal of Clinical Neuroscience, 15,* 388-396.

Bestgen, Y., & Vincze, N. (2012). Checking and bootstrapping lexical norms by means of word similarity indexes. *Behavioral Research Methods*. Advance online publication. doi 10.3758/s13428-012-0195-z

Bertels, J., Kolinsky, R., & Morais, J. (2009). Norms of emotional valence, arousal, threatening value and shocking value for 80 spoken French words – Comparison between neutral and emotional tones of voice. *Psychologica Belgica, 49,* 19-40.

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., 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. *L'Année Psychologique, 104*, 655-694.

Bower, G. H., Gilligan, S. G., & Monteiro K. P. (1981). Selectivity of learning caused by affective states. *Journal of Experimental Psychology: General,* 110, 451-473.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavioral Therapy and Experimental Psychiatry, 25*, 49-59.

Bradley, M. M., & Lang, P.J. (1999a). Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings. Technical report C-1.Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.

Bradley, M. M., & Lang, P. J. (1999b). International affective digitized sounds (IADS): Stimuli, instruction manual and affective ratings. Techincal report B-2. Gainesville, FL: The Center for Research in Psychophysiology, University of Florida

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

Brysbaert, M., Lange, M., & Van Wijnendaele, I. (2000). The effects of age-of-acquisition and frequency-of-occurrence in visual word recognition: Further evidence from the Dutch language. *European Journal of Cognitive Psychology, 12*, 65-85.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41,* 977-990.

Citron, F., Weekes, B., & Ferstl, E. (2009). Evaluation of lexical and semantic features for English emotion words. In K. Alter, M. Horle, M. Lindgren, M. Roll, & J. von Koss Torkildsen (Eds.), *Brain talk: Discourse with and in the brain* (pp.11-20). Lund, Sweden: Media-Tryck.

Compton, R. J. (2003). The interface between emotion and attention: A review of evidence from psychology and neuroscience. *Behavioral and Cognitive Neuroscience Reviews, 2,* 115-129.

Cortese, M. J., & Khanna, M. M. (2007). Age of acquisition predicts naming and lexical-decision performance above and beyond 22 other predictor variables: An analysis of 2,342 words. *The Quarterly Journal of Experimental Psychology, 60,* 1072-1082.

Corson, Y., & Quistrebert, V. (2000). Emotional judgment norms concerning the meanings of ambiguous words. *Current Psychology of Cognition, 19,* 333-392.

Cuthbert, B. N., Bradley, M. M., & Lang, P. J. (1996). Probing picture perception: Activation and emotion. *Psychophysiology, 33,* 103-111.

Dan Glauser, E. S., & Scherer, K. R. (2008). Neuronal processes involved in subjective feeling emergence: Oscillatory activity during an emotional monitoring task. *Brain Topography, 20*, 224-231.

De Moor, W., Ghyselinck, M., & Brysbaert, M. (2000). A Validation study of the Age-of-Acquisition norms collected by Ghyselinck, De Moor, & Brysbaert. *Psychologica Belgica, 40*, 99-114.

Drake, R., & Myers, L. (2006). Visual attention, emotion, and action tendency: Feeling active or passive, *Cognition and Emotion, 20*, 608-622.

Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith, *Handbook of affective sciences* (pp. 572-595). Oxford, UK: Oxford University Press.

Eilola, T. M., & Havelka J. (2010). Affective norms for 210 British English and Finnish nouns. *Behavior, Research, and Methods*, *42*, 134-140.

Ferrand, L., Bonin, P., Méot, A., Augustinova, M., New, B., Pallier, C., & Brysbaert, M. (2008). Age of acquisition and subjective frequency estimates for all generally known monosyllabic French words and their relation with other psycholinguistic variables. *Behavior Research Methods, 40,* 1049-1054.

Fontaine, J. R. J., Poortinga, Y. H., Setiadi, B., & Suprapti, S. M. (2002). Cognitive structure of emotion terms in Indonesia and The Netherlands. *Cognition and Emotion,16*, 61-86.

Fontaine, J., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). The world of emotions is not two-dimensional. *Psychological Science, 18*, 1050–1057.

Frijda, N. H., Kuipers, P., & ter Schure, E. (1989). Relations among emotion, appraisal, and emotional action readiness. *Journal of Personality and Social Psychology, 57*, 212-228.

Ghyselinck, M., Custers, R., & Brysbaert, M. (2003). Age-of-acquisition rations for 2332 Dutch words from 49 different semantic categories. *Psychologica Belgica*, *43*, 181-214.

Ghyselinck, M., De Moor, W., & Brysbaert, M. (2000). Age of acquisition ratings for 2816 Dutch four- and five-letter words. *Psychologica Belgica, 40*, 77-98.

Grühn, D., & Smith, J. (2008). Characteristics for 200 words rated by young and older adults: Age-dependent evaluations of German adjectives (AGE*). Behavior Research Methods, 40*, 1088-1097.

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

Hermans, D., & De Houwer, J. (1994). Affective and subjective familiarity ratings of 740 Dutch words. *Psychologica Belgica, 34* , 115-139.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., &Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin,136*, 390-421.

Kanske, P., & Kotz, S. A. (2010). Leipzig affective norms for German: A reliability study. *Behavior Research Methods, 42*, 987-991.

Keltner, D., Gruenfeld, D. H., & Anderson, C. (2003). Power, approach, and inhibition. *Psychological Review, 110*, 265-284.

Kensinger, E. A. (2004). Remembering emotional experiences: The contribution of valence and arousal. *Reviews in the Neurosciences, 15*, 241-251 .

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new frequency measure for Dutch words based on film subtitles. *Behavior Research Methods, 42*(3), 643-650.

Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 Dutch mono- and disyllabic words and nonwords. *Frontiers in Psychology 1*, 174.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (in press). Age-of-acquisition ratings for 30 thousand English words. *Behavior Research Methods*.

Lahl, O., Göritz, A. S., Pietrowsky, R., & Rosenberg, J. (2009). Using the World-Wide Web to obtain large-scaleword norms: 190,212 ratings on a set of 2,654 German nouns. *Behavior Research Methods, 41,* 13-19.

Laney, C., Campbell, H. V., Heuer, F., & Reisberg, D. (2004). Memory for thematically arousing events. *Memory and Cognition, 32,* 1149-1159.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1998). Emotion, motivation, and anxiety: Brain mechanisms and psychophysiology. *Biological Psychiatry*, *44*, 1248-1263.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8. Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.

Levine L. J., & Pizarro D.A. (2004). Emotion and memory research: A grumpy overview. *Social Cognition, 22,* 530-554.

Messina, D., Morais, J., & Cantraine, F. (1989). Valeur affective de 904 mots de la langue française. *Cahiers de Psychologie Cognitive, 9,* 165-187.

Moors, A. (2010). Automatic constructive appraisal as a candidate cause of emotion. *Emotion Review*, *2,* 139-156.

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition norms for a large set ofobject names and their relation to adult estimates and other variables. *The Quarterly Journal of Experimental Psychology, 50A,* 528-559.

Niedenthal, P. M., Auxiette, C., Nugier, A., Dalle, N., Bonin, P., & Fayol, M. (2004). A prototype analysis of the French category ‘émotion’. *Cognition and Emotion, 18*, 289-312.

Rouckhout, D., & Schacht, R. (2000). Ontwikkeling van een Nederlandstalig interpersoonlijk circumplex. *Diagnostiekwijzer, 3,* 96-118.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana, IL: University of Illinois Press.

Phan, K. L., Wager T., Taylor, S.F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage, 16,* 331-348.

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.

Roseman, I. J., & Evdokas, A. (2004). Appraisals cause experienced emotions: Experimental evidence. *Cognition and Emotion, 18,* 1-28.

Russell, J. A., Bachorowski, J. A., & Fernandez-Dols, J. M. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology, 54,* 329-349.

Sander, D., Grandjean, D., & Scherer, K. R. (2005). A systems approach to appraisal mechanisms in emotion. *Neural Networks, 18,* 317-352.

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion, 23,* 1307-1351.

Smith, P. K., & Bargh, J. A. (2008). Nonconscious effects of power on basic approach and avoidance tendencies. *Social Cognition, 26,* 1-24.

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.

Synoniemen.net. Retrieved March, 21, 2011, from http://synoniemen.net/

Syssau, A., & Font, N. (2005). Evaluations des caractéristiques émotionnelles d'un corpus de 604 mots. *Bulletin de Psychologie, 58*, 361-367.

Van Loon-Vervoorn, W. A. (1989). *Eigenschappen van basiswoorden*. Lisse, The Netherlands: Swets and Zeitlinger.

Võ, M. L.-H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods, 41,* 534-538.

Vohs, K. D., Baumeister, R. F., & Loewenstein, G. (Eds.) (2007). *Do emotions help or hurt decision making?* New York: Russell Sage Foundation.

Williams, J. M. G., Mathews, A., MacLeod, C. (1996). The Emotional Stroop Task and Psychopathology. *Psychological Bulletin, 120,* 3-24.

Table 1

Summary of Variables Included in the Word List with Means (M), Standard Deviations (SD), and Range

Dimension	*M*	*SD*	Range
Valence	3.94	1.06	1.14-6.61
Arousal	4.09	0.84	1.45-6.52
Dominance	4.14	0.70	1.91-6.16
AoA	7.34	1.93	2.03-14.58
Length	6.11	2.30	2.00-18.00
Frequency (log10pm)	0.67	0.89	-1.64-4.26

Table 2

Split-half Reliabilities for Each Variable Within and Between Samples

	Gent	Leuven	Leiden-Amsterdam	Rotterdam
Valence				
Gent	*0.97*			
Leuven	0.97	*0.97*		
Leiden-Amsterdam	0.96	0.96	*0.96*	
Rotterdam	0.95	0.95	0.96	*0.96*
Arousal				
Gent	*0.93*			
Leuven	0.92	*0.92*		
Leiden-Amsterdam	0.89	0.90	*0.91*	
Rotterdam	0.87	0.88	0.90	*0.87*
Dominance				
Gent	*0.87*			
Leuven	0.87	*0.83*		
Leiden-Amsterdam	0.86	0.88	*0.87*	
Rotterdam	0.82	0.82	0.84	*0.82*
AoA				
Gent	*0.96*	-	-	
Rotterdam	0.89	-	-	*0.93*

Note: The split-half reliabilities within samples are based on smaller halfs than those between samples, which may explain why the former are sometimes smaller than the latter.

Table 3

Correlations between the Variables

	Dominance	Arousal	AoA	Freq(log10)	Length
Valence	0.27***	-0.01	-0.17***	0.15***	-0.08***
Dominance		0.59***	0.08***	0.17***	0.08***
Arousal			0.03	0.10***	0.19***
AoA				-0.60***	0.33***
Freq (log10)					-0.25***

***p < .001

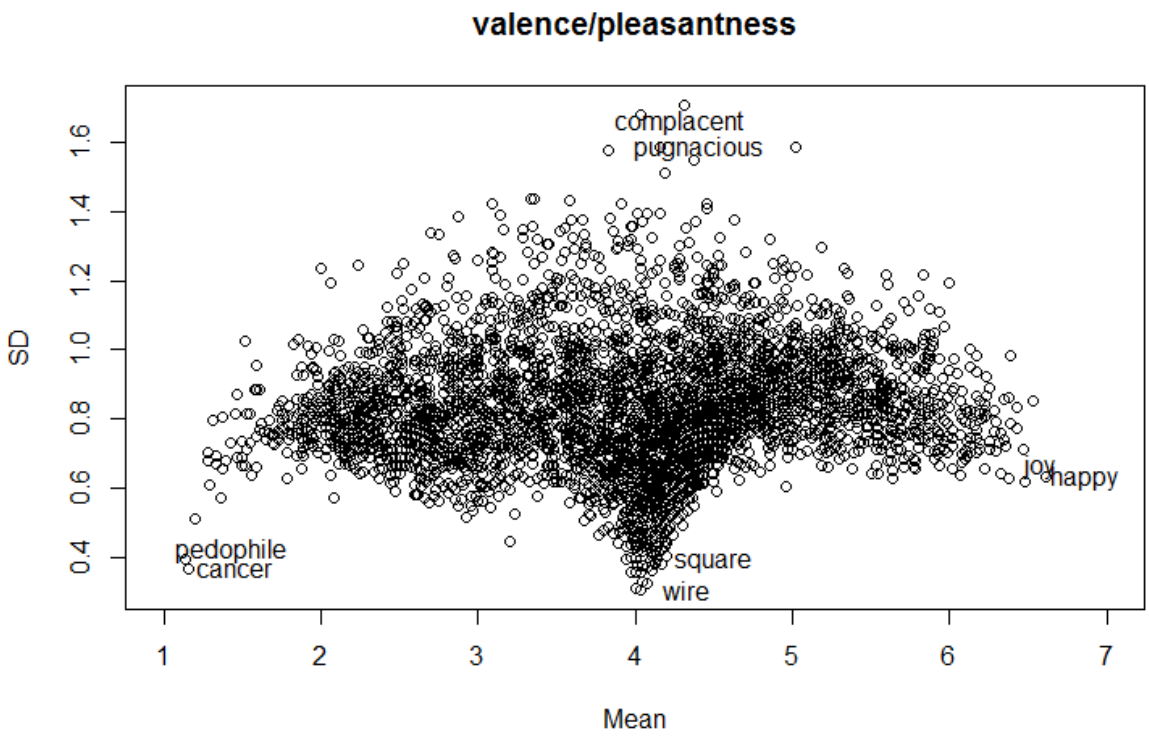

Figure 1. Mean valence/pleasantness ratings plotted against the SDs for these ratings for all 4299 words.

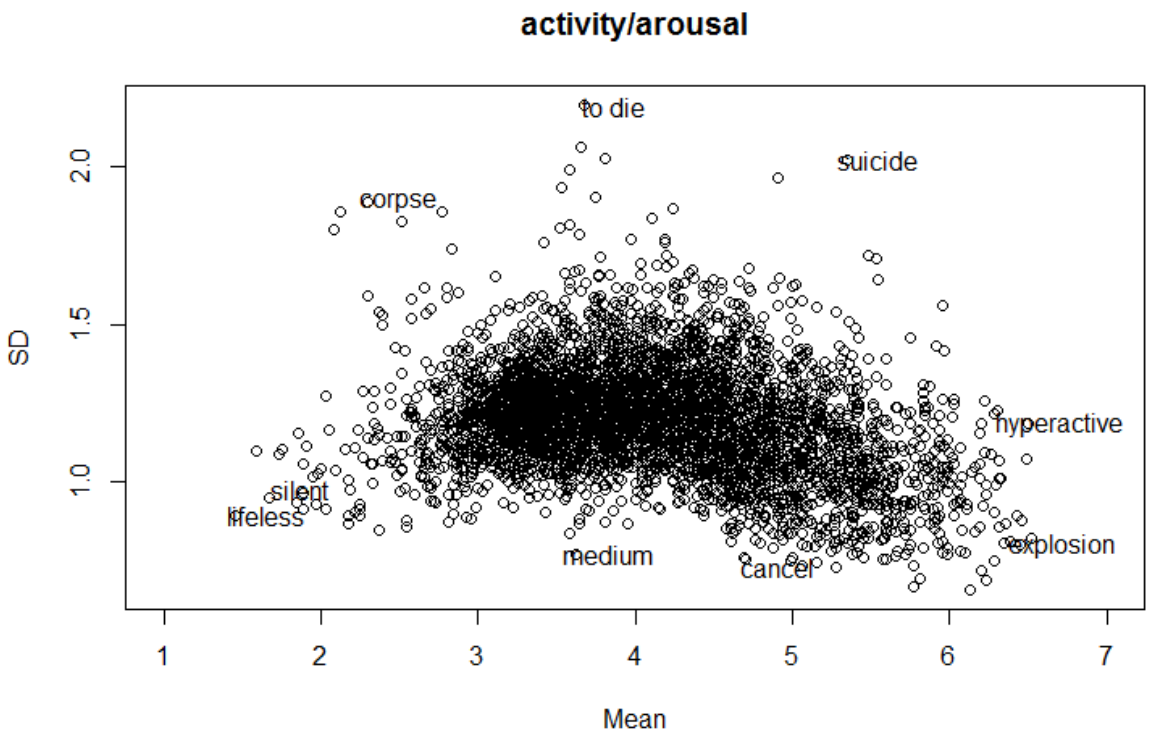

Figure 2. Mean arousal/activity ratings plotted against the SDs for these ratings for all 4299 words.

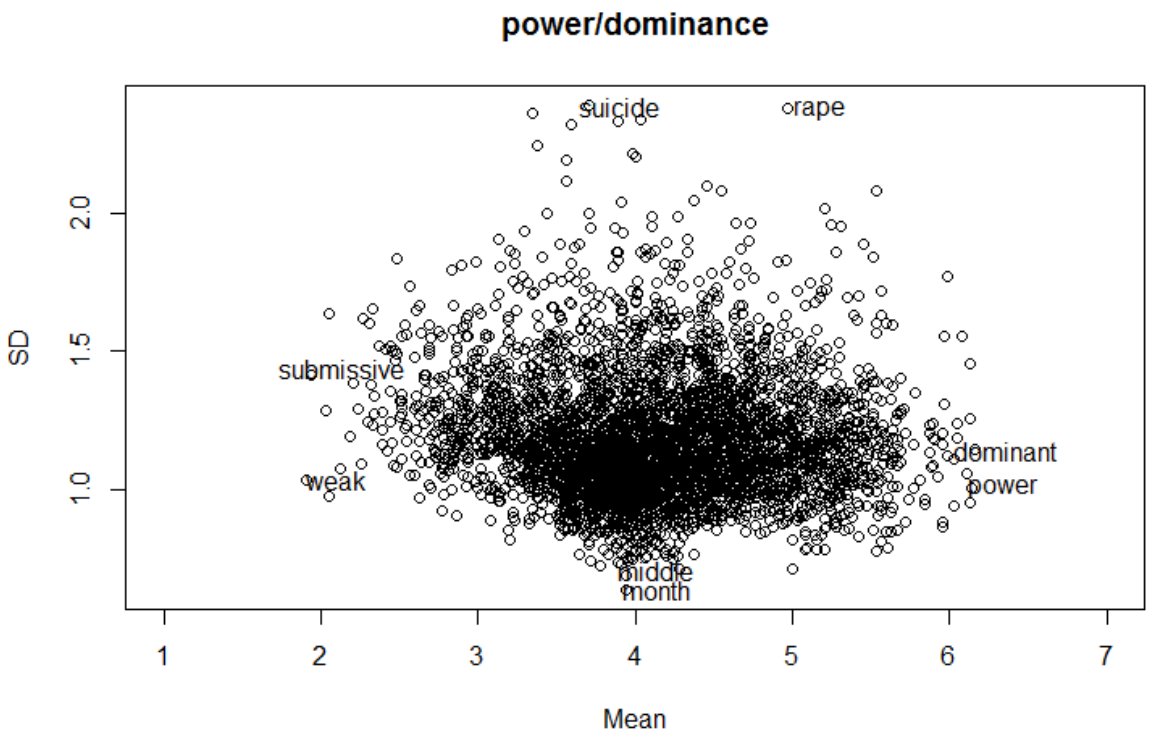

Figure 3. Mean power/dominance ratings plotted against the SDs for these ratings for all 4299 words.

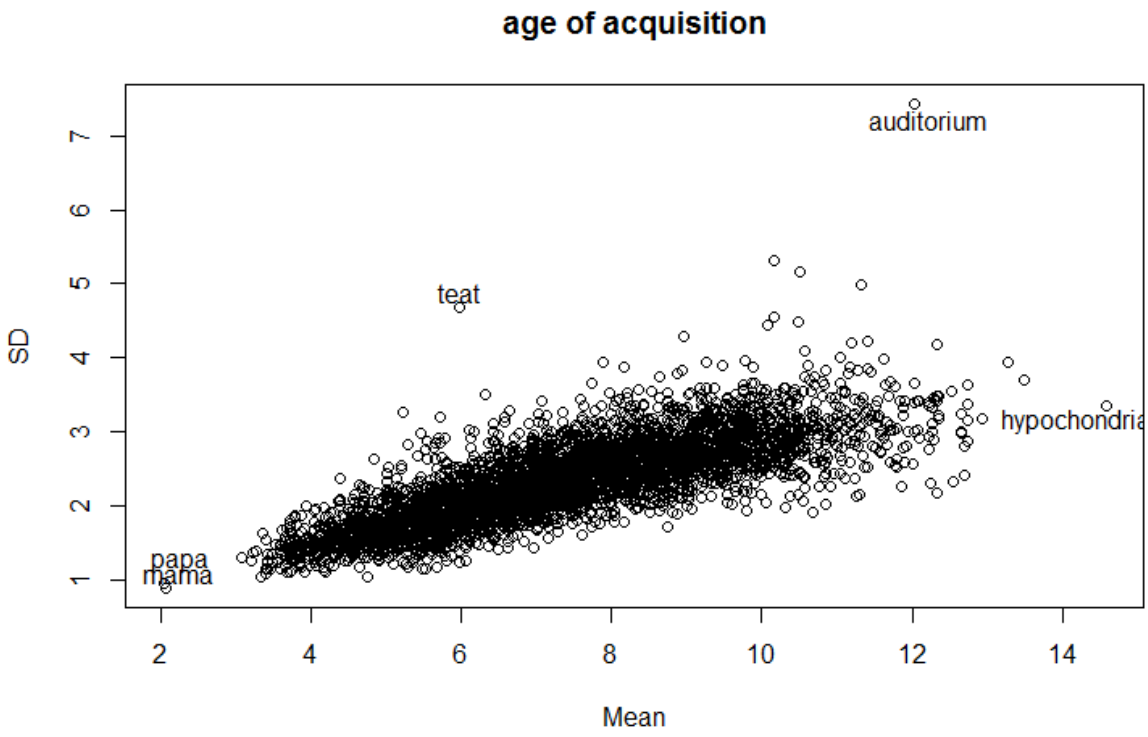

Figure 4. Mean AoA ratings plotted against the SDs for these ratings for all 4299 words.

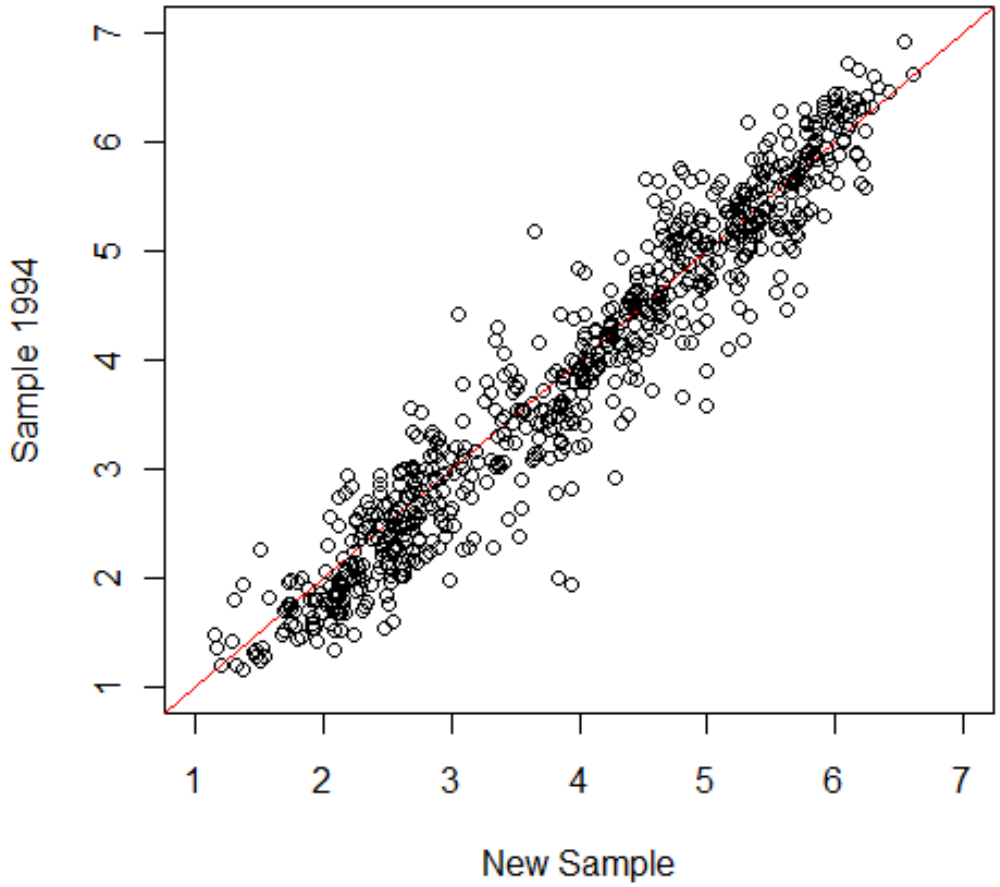

Figure 5. Mean valence ratings in the present sample plotted against the mean valence ratings of the sample of Hermans and De Houwer (1994) for all 715 retested words.

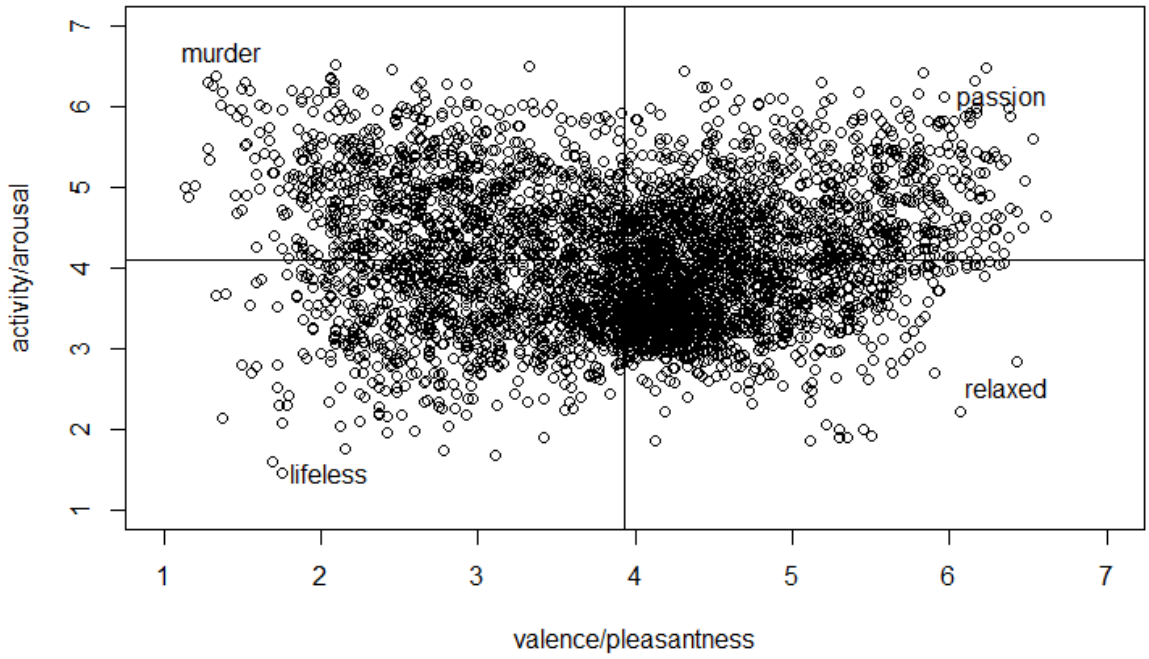

Figure 6. Mean valence/pleasantness ratings plotted against mean arousal/activity ratings.

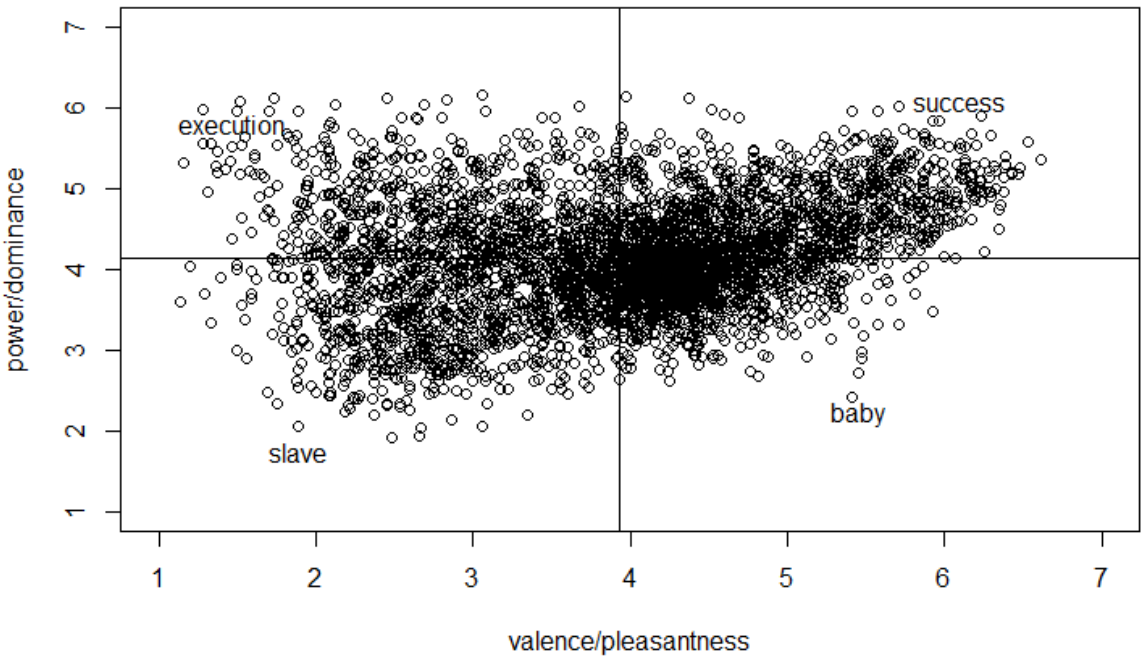

Figure 7. Mean valence/pleasantness ratings plotted against power/dominance ratings.

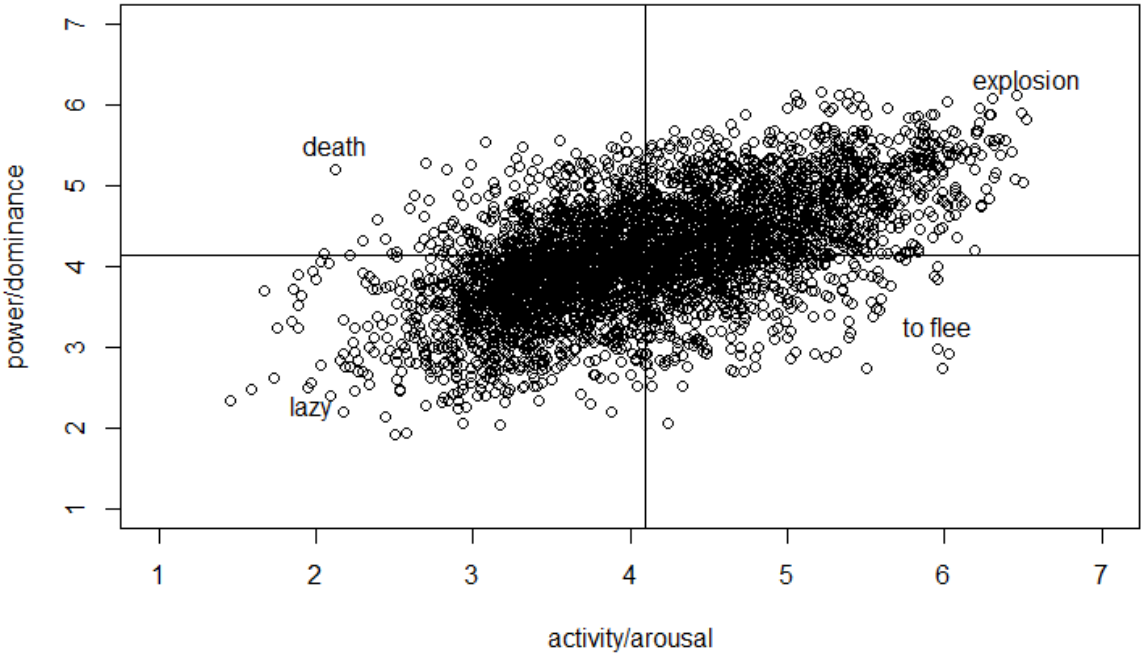

Figure 8. Mean arousal/acitivity ratings plotted against mean power/dominance ratings.