



Current psychology letters

Behaviour, brain & cognition

Vol. 25, Issue 1, 2009 | 2009

Varia

Validation of the Van Overschelde et al. (2004) category norms: Results from five experiments

Ben BAUER and Joanne E. GOURGOUVELIS



Electronic version

URL: <http://journals.openedition.org/cpl/4802>

DOI: 10.4000/cpl.4802

ISSN: 1379-6100

Publisher

Centre PsyCLÉ

Electronic reference

Ben BAUER and Joanne E. GOURGOUVELIS, « Validation of the Van Overschelde et al. (2004) category norms: Results from five experiments », *Current psychology letters* [Online], Vol. 25, Issue 1, 2009 | 2009, Online since 17 March 2009, connection on 08 September 2020. URL : <http://journals.openedition.org/cpl/4802> ; DOI : <https://doi.org/10.4000/cpl.4802>

This text was automatically generated on 8 September 2020.

© All rights reserved

Validation of the Van Overschelde et al. (2004) category norms: Results from five experiments

Ben BAUER and Joanne E. GOURGOUVELIS

- 1 Objects that are judged as good or typical members of a category or that have demonstrated a high probability of being named as members when individuals are prompted with the category, appear to be processed preferentially when presented in a sentence verification task. For example, when asked assign a truth value to a sentence of the form: "A X is a Y", reaction times (RT) for true sentences are shorter when X is a more typical instance in the category Y than for atypical examples. "A robin is a bird" is verified faster than "A bat is a bird" and this effect obtains when the exemplars are presented as pictures rather than words (see, for example, Chumbley, 1986; Freedman & Loftus, 1971; Glass, Holyoak, & Kiger, 1979; Holmes & Ellis, 2006; Kiran & Thompson, 2003; Larochelle & Pineau, 1994; Malt & Smith, 1984; Loftus, 1973; McCloskey & Glucksberg, 1978; McFarland, Duncan & Kellas, 1978; McFarland, Kellas, Klueger, & Juola, 1974). The probability of being named as a member of a prompted category is called instance dominance (see Freedman & Loftus, 1971; Loftus 1973) and this property has been shown to be psychologically relevant by, for example, Loftus (1973) and Smith (1967) who found a benefit on the order of 60-150 ms for category verification in going from low to high instance dominance exemplars. Fundamental to these sorts of studies is the availability of an appropriate set of typicality and instance-dominance norms values for a variety of categories and exemplars.
- 2 Previous research has benefited immensely from the norms provided by Cohen, Bousfield, and Whitmarsh, (1957) and Battig and Montague (1969). Van Overschelde, Rawson, and Dunlosky (2004) report that the Battig and Montague norms have been used in over 1600 studies. Language changes with time; new terminology is born and some falls into disuse or changes in sense. This may render these norms less valid as they age. For example, whereas the waltz and frug were the most popular dances among college undergraduates in 1965, ballet and the tango were the most popular in

2002 (Van Overschelde, Rawson & Dunlosky, 2004). One does not expect a high instance-dominance for the frug if norms were taken on current university-age individuals. Fortunately, Van Overschelde et al., (2004) have performed a great service to the research community in updating and extending the Battig and Montague (1969) the category/exemplar norm set.

- 3 Determination of instance-dominance norms involves supplying category names to a large number of individuals and computing statistics based on the instances (exemplars) they produce for the categories. Van Overschelde et al., gathered such responses from over 600 members of university communities in 3 different states. They computed several statistics on the responses, the most important of which (for the present purposes) is their "TOTAL" value defined as "the proportion of all participants who gave the particular response" (p.291), and is very much like the dominance construct articulated by Freedman and Loftus (1971): "defined as the likelihood that a particular response will be given when S [the subject] is asked to name words that fit in a particular category" (p. 112) and more specifically as instance-dominance, "the percentage of Ss who gave an instance to a particular category" in Loftus (1973, p.70). The norms provided by Van Overschelde et al., range from 1.0 (all respondents generated a given exemplar) down to 0.05 as table in Van Overschelde et al. (2004), but are available for TOTAL down to 0.02 (see, Van Overschelde et al. 2006).
- 4 In the following 5 experiments we correlate the "TOTAL" values from column two of Van Overschelde et al.'s Appendix (hereafter TOTAL) with button-press Reaction Times (RT) from a computer-based category/exemplar verification task using a total of 236 university students. More precisely, the natural logarithm of RT [i.e., $\ln(\text{RT})$] see Rips, Shoben & Smith, 1974; Wilkins, 1971] was correlated with TOTAL. Given the precedents in earlier research, we expected the magnitude of the RT effect from high to low TOTAL to be on the order of 25-150 ms as found by Loftus (1973), Smith (1967), and Glass, Holyoak, and Kiger, (1979). The correlation between TOTAL and $\ln(\text{RT})$ should be negative in all cases; RT should decrease as TOTAL increases. We also expect an average base RT for correct TRUE responses to be 700-1200 ms similar to Loftus (1973) and Smith (1967) and that the probability of a correct "True" response (i.e., a HIT) should increase with TOTAL.
- 5 The 5 experiments were highly similar in implementation with minor modifications in the presentation parameters to establish that the magnitudes of the correlations and effects of TOTAL are not highly influenced by arbitrary display choices. In each experiment 50% of the trials contained TRUE or valid combination of category/exemplar while the other 50% contained invalid or FALSE nonexemplar foils. The participants were students from a population of undergraduates at a small satellite university campus in southern Ontario Canada.

Experiment 1

Method

- 6 *Participants.* Seventy-six undergraduate psychology students (15 male, 61 females) participated. Students were either paid or received course credit for their participation.

- 7 *Stimuli.* Of the 70 categories provided by Van Overschelde et al., 11 were included in this experiment. These categories were selected because they were (intuitively) likely to be culturally appropriate for the student sample tested in the present experiments based on the normative group polled by Van Overschelde et al. For example, categories such as GEMSTONES and FRUITS are more likely to be culturally consistent across these two groups than are COLLEGE/UNIVERSITY FOOTBALL TEAMS, or even STATES.
- 8 For each of the categories, 10 exemplars were selected; 5 from amongst the highest TOTAL and 5 amongst lowest TOTAL. In addition 10 distractor foils were created for each category so that the foils were roughly equated with the exemplars on number of letters and syllables (see Appendix A) for the 11 categories and 20 exemplars/foils per category. Several small modifications to 4 categories were introduced. Parsnip was added to category 43 (VEGETABLES) and assigned a TOTAL of .02, earwig was added to category 45 (INSECT) and assigned a TOTAL .02, ginger was added to category 70 (HERBS) and assigned a TOTAL of .02. Of the three, only ginger appears in the extended tables (containing TOTAL values down to 0.02) available from Van Overschelde et al. where its TOTAL is listed as 0.04. Category names were reduced to a single word, for example category 1 from Van Overschelde et al. "A PRECIOUS STONE" was relabeled "GEMSTONE". This was done to keep the amount of reading to a minimum because RTs were being recorded. Stimuli were sentences presented in the form of "(An) exemplar is (an) category" presented on a laboratory computer. Concrete nouns were preceded by an indefinite article of 'a' or 'an', abstract nouns were not preceded by an indefinite article, according to customary usage in local language usage area (Southern Ontario, Canada).
- 9 The order of presentation of the 220 unique sentences (11 categories x 20 exemplars/foils) was randomized for each participant. The sentences were displayed one at a time in a white, 24pt Helvetica font on a black background using 17" RGB monitors driven by Apple PowerPC computers. The monitors were set to a resolution of 832x624 pixels at 75Hz vertical refresh. The software to control the display and record responses was written using the Video ToolBox routines (Pelli, 1997). Responses were gathered using Apple ADB keyboard using the left arrow key for TRUE responses and the right arrow key for FALSE responses. Each trial began with a 100 ms grey fixation cross positioned at screen position 296, 328 pixels. Upon erasure of the fixation cross the sentence appeared left-justified just to the right of the location where the fixation cross had been. The sentence remained on the screen until a response was made. There was a 1200 ms blank screen between trials. Viewing distance was standardized at 50 cm.
- 10 *Procedure.* The testing room contained 5 workstations and participants were tested in groups of 1 to 5. Once the participant(s) were seated in front of the computer(s), they were instructed that the task was to determine whether sentences displayed one-at-a-time were true or false. An example of a true statement was provided "Mars is a planet" as was a false example "A dog is an ocean". They were instructed to respond as quickly as possible using the arrow keys (which had been marked T and F) while keeping errors to a minimum. No practice trials were given in any of the present experiments. Following the final trial, a plot of the participant's raw RTs (untrimmed for outliers) as a function of the TOTAL values from the Van Overschelde et al. norms was shown and a brief oral interpretation was offered.

Results

- 11 RTs for correct TRUE responses (HITS) were analyzed after trimming for outliers. RTs less than 350ms and greater than 3000ms were excluded from the analyses. Approximately 6% of the responses for TRUE exemplars were errors. The simple correlation between TOTAL and $\ln(\text{RT})$ was $r(7595) = -0.1872$, $p < .001$ indicating that RTs decreased as the TOTAL value of the TRUE exemplar increased. On average, the slope (of the linear values) indicates a 258 ms RT advantage for verifying high TOTAL versus low TOTAL exemplars. A subsequent analysis on raw untrimmed $\ln(\text{RT})$ produce similar results $r(7803) = -0.1992$, $p < .001$ with a 374 ms increase in RT to verify a TRUE exemplar with a low TOTAL versus high TOTAL. A correlation of TOTAL with HITS coded as '+1' and errors (MISSES) as '-1' was $r(8075) = +0.1293$, $p < .001$ demonstrating that as the TOTAL value increases the probability of making an error decreases.

Discussion

- 12 RT decreases with increasing TOTAL value as expected. The correlation was modest but in practical terms of approximately a quarter of a second benefit in going from low to high TOTAL values demonstrates the psychological relevance of the measure.

Experiment 2

- 13 Experiment 2 replicated Experiment 1 with a different sample of students ($N=46$) and a small display modification. Sentences were broken vertically such that the subject of the sentence was presented above the predicate to reduce head and eye movements during reading. The concern was that the modest correlation in Experiment 1 reflected a floor effect on RT because the semantic processing times were swamped by the need for these movements to complete reading the sentences.

Results

- 14 Data from 1 male subject was excluded due to 25% miskeyed responses. Approximately 7% of the responses for TRUE exemplars were errors. The correlation between TOTAL and trimmed $\ln(\text{RT})$ was $r(4480) = -.1703$, $p < .001$ with a 197ms RT advantage for high TOTAL versus low TOTAL exemplars. A subsequent analysis on the untrimmed $\ln(\text{RT})$ produced similar results $r(4556) = -.169$, $p < .001$ and a 250ms advantage. The correlation of TOTAL with accuracy (as described in Experiment 1) was $r(4838) = .1572$, $p < .001$ demonstrating that as the TOTAL value increases the probability of making an error decreases.

Discussion

- 15 The correlation between TOTAL and $\ln(\text{RT})$ for both the trimmed and untrimmed data are similar to results found in Experiment 1. In Experiment 3, four of the categories used in Experiments 1 and 2 will be reused with 13 additional categories introduced. This will give another data point to help assess whether the slopes obtained so far (197 ms, 258 ms) are representative of additional categories (recall that only 11 of 70

categories used by Van Overschelde et al. categories were tested so far). Another concern was that the low correlation could have resulted from frequent reuse of the categories in Experiment 1 and 2. Each category name was presented 20 times with 10 foils and 10 exemplars. This could potentially suppress the effect of TOTAL value and reduce the correlation by priming the categories and exemplars repeatedly. For example, even though "teal" has a low TOTAL value of .08, having seen the category COLOUR numerous times and perhaps related colours "blue" and "green", or possibly even priming the bird category since "teal" is also a duck species. Smith (1967) found the RT effects of instance dominance were attenuated on the second occurrence of an exemplar. While in none of the present experiments an exemplar was repeated to a participant, the fact that the exemplars are highly related may produce a similar effect. For this reason we only used 1 high and 1 low exemplar per category.

Experiment 3

- 16 The categories and exemplars used in Experiment 3 are tabled in Appendix B. Each category was shown only 4 times, once with a high TOTAL exemplar, once with a low TOTAL exemplar and twice with foils. A slope and correlation similar to those found in previous experiments would demonstrate that the original 11 categories had adequately represented the 70 Van Overschelde et al. categories and that the TOTAL value predicts RT equally well. Thirty-two students (5 males, 27 females) were tested. Given 17 categories (13 new categories and 4 categories used previously) each seen 4 times, a total of 68 unique sentences seen by each participant. The mean TOTAL for the high TOTAL exemplars was .783 (SD = .187) and the low TOTAL exemplar mean was .059 (SD = .021).

Results

- 17 Approximately 5% of the responses were MISSES. The correlation between TOTAL and $\ln(\text{RT})$ was $r(1007) = -0.2015$, $p < .001$ with a 221ms RT advantage for high TOTAL exemplars. The simple correlation between TOTAL and untrimmed $\ln(\text{RT})$ was $r(1030) = -0.1922$, $p < .001$ with 287ms advantage. A correlation of TOTAL with accuracy as in the previous 2 experiments was $r(1055) = .1578$, $p < .001$. A repeated measure ANOVA showed that RT for high TOTAL exemplars (1180ms) was significantly faster than the 1364ms RT for low TOTAL exemplar $F(1,31) = 40.62$ $p < .001$ (MSE = 13326). The 184 ms benefit for high TOTAL exemplars (or the 221.1 slope from the regression analysis) is in the same range as Experiment 1 and 2. Slopes and correlations from all 5 experiments see Tables 1 and 3 in Appendix C.

Experiment 4

- 18 Loftus (1973) found that "category dominance" (empirical probability that a particular category name is given as a response to an exemplar) is a better predictor of RT when the exemplar precedes the category in a verification task. However, "instance dominance" (empirical probability that a particular exemplar name is generated in a response to a category) better predicted RT when the category preceded the exemplar in the task. In that experiment the first item (category or exemplar) was visible for 1

second before the second item (exemplar of category). The TOTAL value tabled in Van Overschelde et al. is an "instance dominance" measure. In Experiments 1-3 both exemplar and category were simultaneously available either in the same sentence presented horizontally across the screen (Experiment 1) or exemplar above category in Experiments 2 and 3. In Experiment 4 the category was presented above the exemplar because the results from Loftus (1973) suggest that instance dominance qua TOTAL has a more potent effect when the category precedes the exemplar (see also Chumbley, 1986). Note that with vertical presentation it is still likely that participants viewed both category and exemplar simultaneously. However reading order would likely have been category then exemplar. Fifty-one undergraduate psychology students were tested using the same categories and exemplars as in Experiments 1 and 2. The response keys were reversed (right arrow key for TRUE, left arrow key for FALSE) to make sure results were not peculiar to mapping of the response keys.

Results

- 19 Approximately 11% of the responses for TRUE exemplars were errors. The correlation between TOTAL and $\ln(RT)$ was $r(4937) = -0.2359$, $p < .001$ with a 263ms advantage for verifying high TOTAL versus low TOTAL exemplars. A subsequent analysis on the raw untrimmed log RTs produce similar results $r(4983) = -0.2166$, $p < .001$ between RT and TOTAL and a 287ms advantage. A correlation of TOTAL with accuracy as in Experiments 1-4 was $r(5541) = 0.1341$, $p < .001$ demonstrating that as the TOTAL value increases the probability of making an error decreases. These values are consistent with previous experiments.

Discussion

- 20 Across 4 experiments containing several presentation manipulations, correlations, slopes have been similar. This was also true in Experiment 3 where several new categories were introduced. One potential reason for the modest correlation could be that the nature of the foils used on the FALSE trials interacted with the nature of the exemplars used on the TRUE trials. The foils (Experiments 1-4) were generated to approximately match the exemplars number of letters and syllables. These are only physical characteristics of the words. In Experiment 5, psychological characteristics of the foils were more closely matched to the exemplars (the authors thank James Van Overschelde for this suggestion).

Experiment 5

- 21 To produce a set of foils approximately matched on word frequency, concreteness, familiarity, meaningfulness and number of letters and syllables. Foils were selected according to the values on these variables using the MRC Psycholinguistic Database (www.psy.uwa.edu.au/mrcdatabase/uwamrc.htm). If the word frequency of the foils had affected the slopes and correlations in Experiments 1-4, using this matched set of foils may produce slopes and correlations different from those previous experiments.

Stimuli and Procedure

- 22 The categories and exemplars used were the same as used in Experiments 1, 2, and 4 however the foils for Experiment 5 were selected in the following manner: For each TRUE exemplar, a matching FALSE foil exemplar was created with similar word frequency, concreteness, imageability, familiarity, meaningfulness and similar number of letters/syllables as the TRUE exemplar. In Appendix C, the 10 TRUE exemplars for each category have nonzero TOTAL values while the 10 FALSE exemplars have a TOTAL of zero. Word frequency was based on frequency of occurrence as given in the norms of Kučera and Francis (1967). Printed familiarity (FAM), concreteness (CONC) and imageability (IMAG) values were derived from merging three sets of norms: Paivio (unpublished), Toglia and Battig (1978) and Gilhooly and Logie (1980). Meaningfulness (MEANC) values were based on norms from Toglia and Battig (1978). These measures, and their derivations are described at the MRC Psycholinguistic Database website: www.psy.uwa.edu.au/mrcdatabase/uwamrc.htm.
- 23 See Appendix C for the 11 categories and 20 exemplars per category. See Appendix D for the mean category values for each criterion. Thirty-one students were tested. Words were presented as "exemplar category".

Results

- 24 Approximately 7% of the responses for TRUE exemplars were MISSES. The correlation between TOTAL and $\log(\text{RT})$ was $r(3107) = -0.2081$, $p < .001$ with a 251ms advantage for verifying high TOTAL versus low TOTAL exemplars. Analysis on untrimmed $\log \text{RTs}$ produce similar results $r(3142) = -0.1866$, $p < .001$ with a 276ms advantage. The correlation of TOTAL with accuracy was $r(3358) = +0.1568$, $p < .001$ demonstrating that as the TOTAL value increases the probability of making an error decreases. An additional regression analysis was undertaken to determine the contributions of word frequency (from Kučera & Francis, 1967) and TOTAL in predicting $\ln(\text{RT})$. The simple correlation between word frequency and TOTAL was $r = 0.4151$. A multiple regression analysis with word frequency and TOTAL as predictors showed that TOTAL was the more important predictor. Unique variance accounted for by word frequency was essentially zero when TOTAL was removed first. However, with word frequency removed the beta value for TOTAL was -0.2289 . The simple correlation between TOTAL and $\ln(\text{RT})$ in this analysis was $r = -0.2298$ which is slightly higher than the -0.2081 value stated above because only exemplars for which a Kučera & Francis word frequency value was available (72 out of 110) were included in the multiple regression. Arbitrarily setting the word frequency for these 38 words to 1.0, has no effect on the contribution of word frequency.

General Discussion

- 25 In 5 experiments controlling for psychological and physical characteristics of the stimuli, the TOTAL values provided by Van Overschelde et al. indicate a mean RT benefit of 238 ms in going from low TOTAL to high TOTAL exemplars. The average correlation between the TOTAL value and $\ln(\text{RT})$ was approximately -0.20 across experiments (computed according to the method prescribed in Alexander, 1990) which is relatively modest but consistent across the 5 experiments. No doubt, there are

interesting effects of many other psychological properties of the categories and exemplars that contribute to the RT, slopes and correlations obtained here (see Chumbley, 1986; Hernández-Muñoz, Izura & Ellis, 2006; Larochelle & Pineau, 1994; Loftus & Suppes, 1972; McFarland et al., 1974). However the goal of the present experiments was to demonstrate the effectiveness of the TOTAL variable in predicting RT differences, which it consistently does. Over the set of categories/exemplars used this difference is approximately one quarter of a second which on a base RT of around 1200ms amounts to an approximate 20% change in RT. This is substantial. Data from the four similar experiments (Exp. 1, 2, 4, 5) were combined and analyzed using multiple regression. Criterion scaling (see, Pedhazur, 1977; Schumacker & Williams, 1993) was used to account for the repeated measures within-subjects and within-categories. Additional predictors in the regression were the number of letters, number of phonemes, and number of syllables in the exemplar (according to local pronunciation). The semipartial correlation of $\ln(\text{RT})$ with TOTAL was $r=-0.166$, indicating that the unique variance contribution of TOTAL on $\ln(\text{RT})$ was significant (the semipartial correlations for each experiment were: Exp 1, -0.170; Exp 2, -0.133; Exp 4, -0.187; Exp 5, -0.195).

- 26 There are a few small concerns about the implementation of the presented experiments, none of which are assumed to be severely limiting. At this point the effects of the slight changes in the category prompts (e.g., Van Overschelde et al.'s category "A PRECIOUS STONE" was relabeled "GEMSTONE" in the present experiments) are unknown. In some respect the participants in the present experiments were similar to those used to generate the Van Overschelde et al., norms (e.g. university/college students with likely the same age range and first language). At the same time students in the present experiment were mostly psychology majors or minors on a small satellite campus of a Canadian university, admittedly of the 236 participants in these 5 experiments the male to female ratio was 1:4 which could influence correlations and slopes.
- 27 Of the primary 11 categories used the correlation between RT and TOTAL ranges from -0.08 to -0.38. Thus power of the TOTAL value to predict RT varies across categories. For example, excluding Experiment 4, the magnitude of the correlation in the COLOUR and TOOL categories were consistently ranked first or second, whereas the FLOWER and HERB categories at best 8 out of 11. These relative rankings maybe in part driven by the male/female ratio or geographic locality. In Experiment 4, the correlation and slopes are based on only 2 TOTAL values (one high and one low) per category thus little faith is given in their category correlations, nevertheless as a group the correlations were quite similar to the correlations in the other experiments (see Table 1 for category wise correlations and slopes for experiments and Table 2 in Appendix C for the same values ranked within a category). Looking at slope values in Table 1 (in Appendix C) some category RTs are independent of the Van Overschelde et al. values and in others the benefit is large, from essentially no benefit up to 449 ms. To increase the ability to predict RT, the TOTAL value could be used with the other physical and psychological properties (e.g., age of acquisition: see, Holmes & Ellis, 2006; Morrison, Chappell, Ellis, 1997, context variability: see, Steyvers & Malmberg, 2003, experienced co-occurrence of words: Griffiths, Steyvers, Tenenbaum, 2007 and many others). Perhaps other tasks or paradigms such as priming, recall or recognition memory, verification tasks with deadline RT or degraded stimuli, would produce quite different correlations and slopes.

The expected magnitude of the RT effect in going from low to high TOTAL was as big as or larger than that found in previous studies that had used Battig and Montague (1969) or Cohen, Bousfield, and Whitmarsh (1957) norms. As Van Overschelde et al. (2004) had noted those norms had been critical in over 1600 papers but are now severely dated. The research community will surely find this new set of norms to be highly useful contribution in conducting research.

BIBLIOGRAPHY

- Alexander, R. A. (1990). A note on averaging correlations. *Bulletin of the Psychonomic Society*, 28, 335-336.
- Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut norms. *Journal of Experimental Psychology*, 80, 1-46.
- Chumbley, J. I. (1986). The roles of typicality, instance dominance, and category dominance in verifying category membership. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 257-267.
- Cohen, B. H., Bousfield, W. A., & Whitmarsh, G. A. (1957). Cultural norms for verbal items in 43 categories (Tech. Rep. No. 22). Storrs, CT: University of Connecticut.
- Freedman, J. L., & Loftus, E. F. (1971). Retrieval of words from long-term memory. *Journal of Verbal Learning and Verbal Behaviour*, 10, 107-115.
- Glass, A. L., Holyoak, K. J., & Kiger, J. I. (1979). Role of antonymy relations in semantic judgments. *Journal of Experimental Psychology: Human, Learning and Memory*, 5, 598-606.
- Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, 114, 211-244. cocosci.berkeley.edu/tom/papers/topicsreview.pdf
- Hernández-Muñoz, N., Izura, C., & Ellis, A. W. (2006). Cognitive aspects of lexical availability. *European Journal of Cognitive Psychology*, 18, 730-755.
- Holmes, S. J., & Ellis, A. W. (2006). Age of acquisition and typicality effects in three object processing tasks. *Visual Cognition*, 13, 884-910.
- Kiran, S., & Thompson, C. K. (2003). Effect of typicality on online category verification of animate category exemplars in aphasia. *Brain and Language*, 85, 441-450. <https://webpace.utexas.edu/kirans/www/RT%20paper.pdf>
- Larochelle, S., & Pineau, H. (1994). Determinants of response times in the semantic verification task. *Journal of Memory and Language*, 33, 796-823.
- Loftus, E. F. (1973). Category dominance, instance dominance, and categorization time. *Journal of Experimental Psychology*, 97, 70-74.
- Loftus, E. F., & Suppes, P. (1972). Structural variables that determine the speed of retrieving words from long-term memory. *Journal of Verbal Learning and Verbal Behaviour*, 11, 770-777.

- Malt, B. C., & Smith, E. E. (1984). Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behavior*, 23, 250-269.
- McCloskey, M. E. & Glucksberg, S. (1978). Natural categories: well defined or fuzzy sets? *Memory & Cognition*, 6, 462-472.
- McFarland, C. E., Duncan, E. M., & Kellas, G. (1978). Isolating the typicality effect in semantic memory. *Quarterly Journal of Experimental Psychology*, 30, 251-262.
- McFarland, C. E., Kellas, G., Klueger, K., & Juola, J. F. (1974). Category similarity, instance dominance and categorization time. *Journal of Verbal Learning and Verbal Behavior*, 13, 698-708.
- Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *Quarterly Journal of Experimental Psychology*, 50A, 528-559.
- MRC Psycholinguistic Database: Machine Usable Dictionary. Version 2.00 www.psy.uwa.edu.au/mrcdatabase/uwamrc.htm
- Pedhazur, E.J., (1977). Coding subjects in repeated measures designs. *Psychological Bulletin*, 84, 298-305.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, 10, 437-42. <http://psych.nyu.edu/pelli/pubs/pelli1997videotoolbox.pdf>
- Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behaviour*, 12, 1-20.
- Schumacker, R.E., & Williams, J.D. (1993). Teaching ordinal and criterion scaling in multiple regression. *Multiple Linear Regression Viewpoints*, 20, 25-31.
- Smith, E. E. (1967). Effects of familiarity on stimulus recognition and categorization. *Journal of Experimental Psychology*, 3, 324-332.
- Steyvers, M., & Malmberg, K. J. (2003). The effect of normative context variability on recognition memory. *Journal of Experimental Psychology*, 5, 760-766. <http://psiexp.ss.uci.edu/research/papers/SteyversMalmberg2003.pdf>
- Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An update and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language*, 50, 289-335.
- Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2006). Addendum to "Category norms: An updated and expanded version of the Battig and Montague (1969) norms". *Journal of Memory and Language*, 54, 633.
- Wilkins, A. J. (1971). Conjoint frequency, category size, and categorization time. *Journal of Verbal Learning & Verbal Behavior*. 10, 4, 382-385.

ABSTRACTS

Van Overschelde, Rawson, and Dunlosky (2004) collected category/exemplar norms from over 600 participants at three American universities. For each of 70 categories, participants generated as many exemplars as possible in 30 seconds. One measure computed from their data was the "TOTAL" statistic - the proportion of participants that generated a particular exemplar given a category. Five experiments in the present investigation tested the ability of the TOTAL statistic to

predict Reaction Time (RT) in a category/exemplar verification task with 236 participants. The simple correlation between "TOTAL" and the natural log of RT was approximately -0.20 with the average cost (slope) from high TOTAL (near 1.0) to low TOTAL (0.02) on the order of 250 ms. The results are similar in magnitude to previous research using older Battig and Montague norms which the Van Overschelde et al. norms supersede suggesting that the updated norms are a suitable contemporary replacement.

Van Overschelde, Rawson et Dunlosky (2004) ont rassemblé des normes catégorielles recueillies auprès de plus de 600 participants provenant de trois universités américaines. Pour chacune des 70 catégories, les participants ont produit le plus d'éléments/de réponses possible en 30 secondes. La statistique TOTALE est l'une des mesures calculées à partir de leurs données. Il s'agit de la proportion de participants qui ont produit un élément/réponse en particulier pour une catégorie donnée. Cinq expériences menées dans la présente étude ont porté sur la capacité de la statistique TOTALE de prédire le temps de réaction (TR) pour une tâche de vérification catégorielle auprès de 236 participants. La corrélation simple entre le TOTAL et le logarithme népérien de TR était d'environ -0,20, le coût moyen (pente) allant du TOTAL élevé (près de 1,0) au TOTAL bas (0,02) étant de l'ordre de 250 ms. Les résultats sont de même grandeur que ceux des recherches précédentes reposant sur les normes plus anciennes de Battig et Montague, ce qui tend à indiquer en l'occurrence que les normes de Van Overschelde et al. peuvent aujourd'hui remplacer avantageusement les anciennes normes.

INDEX

Keywords: category, exemplar, instance dominance, sentence verification, Word frequency

AUTHORS

BEN BAUER

Trent University Oshawa

JOANNE E. GOURGOUVELIS

Trent University Oshawa