1   Investigating the Interaction between Associative, Semantic, and Thematic Database Norms

2   for Memory Judgments and Retrieval

3   Nicholas P. Maxwell[1] & Erin M. Buchanan[1]

4   [1] Missouri State University

5   Author Note

6   Nicholas P. Maxwell is a graduate student at Missouri State University. Erin M.

7   Buchanan is an Associate Professor of Psychology at Missouri State University.

8   Correspondence concerning this article should be addressed to Nicholas P. Maxwell,

9   901 S. National Ave, Springfield, MO, 65897. E-mail: maxwell270@live.missouristate.edu

₁₀   Abstract

₁₁ This study examined the interactive relationship between semantic, thematic, and associative

₁₂ word pair strength in the prediction of item judgments and cued-recall performance.

₁₃ Participants were recruited from Amazon's Mechanical Turk and were given word pairs of

₁₄ varying relatedness to judge for their semantic, thematic, and associative strength. After

₁₅ completing a distractor task, participants then completed a cued recall task. First, we sought

₁₆ to expand previous work on memory judgments to include three types of judgments of

₁₇ memory, while also replicating bias and sensitivity findings. Next, we tested for an

₁₈ interaction between the three database norms (FSG, COS, and LSA) when predicting

₁₉ participant judgments and also extended previous work to test for interactions between the

₂₀ three database norms when predicting recall. Significant three-way interactions were found

₂₁ between FSG, COS, and LSA when predicting judgments and recall. For low semantic

₂₂ feature overlap, thematic and associative strength were competitive; as thematic strength

₂₃ increased, associative predictiveness decreased. However, this trend reversed for high

₂₄ semantic feature overlap, wherein thematic and associative strength were complementary as

₂₅ both set of simple slopes increased together. Overall, our findings indicate the degree to

₂₆ which the processing of associative, semantic, and thematic information impacts cognitive

₂₇ processes such as retrieval and item judgments, while also examining the underlying,

₂₈ interactive relationship that exists between these three types of information.

₂₉   *Keywords:* judgments, memory, association, semantics, thematics

Investigating the Interaction between Associative, Semantic, and Thematic Database Norms

for Memory Judgments and Retrieval

The study of cognition has a rich history of exploring the role of association in human memory. One key finding is that elements of cognitive processing play a critical role in how well an individual retains learned information. Throughout the mid-20th century, researchers investigated this notion, particularly through the use of paired-associate learning (PAL). In this paradigm, participants are presented with a pair of items and are asked to make connections between them so that the presentation of one item (the cue) will in turn trigger the recall of the other (the target). Early studies of this nature focused primarily on the effects of meaning and imagery on recall performance. For example, Smythe and Paivio (1968) found that noun imagery played a crucial role in PAL performance; subjects were much more likely to remember word-pairs that were low in meaning similarity if imagery between the two was high. Subsequent studies in this area focused on the effects of mediating variables on PAL tasks as well as the effects of imagery and meaningfulness on associative learning (Richardson, 1998), with modern studies shifting their focus towards a broad range of applied topics such as how PAL is affected by aging (Hertzog, Kidder, Powell-Moman, & Dunlosky, 2002), its impacts on second language acquisition (Chow, 2014), and even in evolutionary psychology (Schwartz & Brothers, 2013).

Early PAL studies routinely relied on stimuli generated from word lists that focused extensively on measures of word frequency, concreteness, meaningfulness, and imagery (Paivio, 1969). However, the word pairs in these lists were typically created due to their apparent relatedness or frequency of occurrence in text. While lab self-generation appears face valid, one finds that this method of selection lacks a decisive method of defining the underlying relationships between the pairs (Buchanan, 2010), as these variables only capture psycholinguistic measurements of an individual concept (i.e., how concrete is *cat* and word occurrence). PAL is, by definition, used on word pairs, which requires examining concept relations in a reliable manner. As a result, free association norms have become a common

means of indexing associative strength between word pairs (Nelson, McEvoy, & Schreiber, 2004).

As we will use several related variables, it is important to first define association as the context based relation between concepts, usually found in text or popular culture (Nelson, McEvoy, & Dennis, 2000). Such word associations typically arise through their co-occurrence together in language. For example, the terms *peanut* and *butter* have become associated over time through their joint use to depict a particular type of food, though separately, the two concepts share very little in terms of meaning. To generate these norms, participants engage in a free association task, in which they are presented with a cue word and are asked to list the first related target word that comes to mind. The probability of producing a given response to a particular cue word, or forward strength, can then be determined by dividing the number of participants who produced the response in question by the total number of responses generated for that word (FSG; Nelson et al., 2000). Using this technique, researchers have developed databases of associative word norms that can be used to generate stimuli with a high degree of reliability. Many of these databases are now readily available online, with the largest one consisting of over 72,000 associates generated from more than 5,000 cue words (Nelson et al., 2004).

Similar to association norms, semantic word norms provide researchers with another option of constructing stimuli for use in word-pair tasks. These norms measure the underlying concepts represented by words and allow researchers to tap into aspects of semantic memory. Semantic memory is best described as an organized collection of our general knowledge and contains information regarding a concept's meaning (Hutchison, 2003). Models of semantic memory broadly fall into one of two categories. Connectionist models (Rogers & McClelland, 2006; e.g, Rumelhart, McClelland, & Group, 1986) portray semantic memory as a system of interconnected units representing concepts, which are linked together by weighted connections representing knowledge. By triggering the input units, activation will then spread throughout the system activating or suppressing connected units

84 based on the weighted strength of the corresponding unit connections (Jones, Willits, &

85 Dennis, 2015). On the other hand, distributional models of semantic memory posit that

86 semantic representations are created through the co-occurrences of words together in a body

87 of text and suggest that words with similar meanings will appear together in similar contexts

88 (Riordan & Jones, 2011). Popular distributional models of semantic memory include Latent

89 Semantic Analysis (LSA; Landauer & Dumais, 1997) and the Hyperspance Analogue to

90 Language (HAL; Lund & Burgess, 1996).

91   Feature production tasks are a common means of producing semantic word norms

92 (Buchanan, Holmes, Teasley, & Hutchison, 2013; McRae, Cree, Seidenberg, & McNorgan,

93 2005; Vinson & Vigliocco, 2008) In such tasks, participants are shown the name of a concept

94 and are asked to list what they believe the concept's most important features to be (McRae

95 et al., 2005). Several statistical measures have been developed which measure the degree of

96 feature overlap between concepts. Similarity between any two concepts can be measured by

97 representing them as vectors and calculating the cosine value (COS) between them (Maki,

98 McKinley, & Thompson, 2004). Cosine values range from 0 (unrelated) to 1 (perfectly

99 related). For example, the pair *hornet - wasp* has a COS of .88, indicating a high degree of

100 overlap between the two concepts. Feature overlap can also be measured by JCN, which

101 involves calculating the information content value of each concept and the lowest

102 super-ordinate shared by each concept using an online dictionary, such as WordNET (Miller,

103 1995). The JCN value is then computed by summing together the difference of each concept

104 and its lowest super-ordinate (Jiang & Conrath, 1997; Maki et al., 2004). The advantage to

105 using COS values over JCN values is the limitation of JCN being tied to a somewhat static

106 dictionary database, while a semantic feature production task can be used on any concept to

107 calculate COS values. However, JCN values are less time consuming to obtain if both

108 concepts are in the database (Buchanan et al., 2013).

109   Semantic relations can be broadly described as being taxonomic or thematic in nature.

110 Whereas taxonomic relationships focus on the connections between features and concepts

within categories (e.g., *bird - pidgeon*), thematic relationships center around the links between concepts and an overarching theme or scenario (e.g., *bird - nest*; Jones & Golonka, 2012). Jouravlev and McRae (2016) provide a list of 100 thematic relatedness production norms, which were generated through a task similar to feature production in which participants were presented with a concept and were asked to list names of other concepts they believed to be related. Distributional models of semantic memory also lend themselves well to the study of thematic word relations. Because these models are text based and score word pair relations in regard to their overall context within a document, they assess thematic knowledge as well as semantic knowledge. Additionally, text based models such as LSA are able to account for both the effects of context and similarity of meaning, bridging the gap between associations and semantics (Landauer, Foltz, & Laham, 1998).

Discussion of these measures then leads to the question of whether each one truly assesses some unique concept or if they simply tap into our overall linguistic knowledge. Taken at face value, word pair associations and semantic word relations appear to be vastly different, yet the line between semantics/associations and thematics is much more blurred. While thematic word relations are indeed an aspect of semantic memory and include word co-occurrence as an integral part of their creation, themes also appear to be indicative of a separate area of linguistic processing. Previous research by Maki and Buchanan (2008) appears to confirm this theory. Using clustering and factor analysis techniques, they analyzed multiple associative, semantic, and text based measures of associative and semantic knowledge. First, their findings suggested associative measures to be separate from semantic measures. Additionally, semantic information derived from lexical measures (e.g., COS, JCN) was found to be separate from measures generated from analysis of text corpora, suggesting that text based measures may be more representative of thematic information.

While it is apparent that these word relation measures are assessing different domains of our linguistic knowledge, care must be taken when building experimental stimuli through the use of normed databases, as many word pairs overlap on multiple types of measurements,

138   and even the first studies on semantic priming used association word norms for stimuli

139   creation (Lucas, 2000; Meyer & Schvaneveldt, 1971; Meyer, Schvaneveldt, & Ruddy, 1975).

140   This observation becomes strikingly apparent when one desires the creation of word pairs

141   related on only one dimension. One particular difficulty faced by researchers comes when

142   attempting to separate association strength from feature overlap, as highly associated items

143   tend to be semantically related as well. Additionally, a lack of association strength between

144   two items may not necessarily be indicative of a total lack of association, as traditional

145   norming tasks typically do not produce a large enough set of responses to capture all

146   available associations between items. Some items with extremely weak associations may

147   inevitably slip through the cracks (Hutchison, 2003). As such, the present study seeks to

148   provide further insight by examining how different levels of associative overlap (measured in

149   FSG), semantic overlap (feature overlap measured with COS), and thematic overlap

150   (measured with LSA) affect cognitive tasks such as short term item retrieval and item

151   relatedness judgments. Instead of focusing solely on one variable or trying to create stimuli

152   that represent only one form of relatedness, we included a range of each of these variables to

153   explore their potential interaction.

154   Specifically, this research was conceptualized within the framework of a three-tiered

155   view of the interconnections between these systems as it relates to processing concept

156   information. The three-tiered view is inspired by models of reading and naming, particularly

157   the triangle models presented by Seidenberg and Mcclelland (1989) and Plaut, McClelland,

158   Seidenberg, and Patterson (1996). These models explored the nature of reading as

159   bidirectional relations between semantics, orthography, and phonology. In this research, we

160   examine if the semantic, associative, and thematic systems are interactive for judgment and

161   recall processes, much like the proposed interactive nature of phonology, orthographics, and

162   semantics for reading and naming processes. Potentially, association, semantic, and thematic

163   facets of word relation each provide unique component that can be judged and used for

164   memory, thus, suggesting three separate networks of independent information. This view

165 seems unlikely, in that research indicates that there is often overlap in the information

166 provided by each measure of word-pair relatedness. Instead, dynamic attractor networks, as

167 proposed by Hopfield (1982) and McLeod, Shallice, and Plaut (2000) may represent better

168 represent the interplay between these representations of concepts, as these models posit a

169 similar feedback relationship between concepts in a network. Using these models as a

170 theoretical framework for our study, we sought to understand how these three types of

171 word-pair information may interact when judgment and recall processes were applied to

172 concept networks.

Application to Judgment Studies

174 Traditional judgment of learning tasks (JOL) can be viewed as an application of the

175 PAL paradigm; participants are given pairs of items and are asked to judge how accurately

176 they would be able to correctly match the target with the cue on a recall task. Judgments

177 are typically made out of 100, with a participant response of 100 indicating full confidence in

178 recall ability. In their 2005 study, Koriat and Bjork examined overconfidence in JOLs by

179 manipulating associative relations (FSG) between word-pairs and found that subjects were

180 more likely to overestimate recall for pairs with little or no associative relatedness.

181 Additionally, this study found that when accounting for associative direction, subjects were

182 more likely to overestimate recall for pairs that were high in backwards strength but low in

183 forward strength. To account for this finding, the authors suggested that JOLs may rely

184 more heavily on overlap between cue and target with the direction of the associative

185 relationship being secondary. Take for example the pair *feather - bird*, which has a FSG of

186 .051 and a BSG of .359. This item pair also has a cosine value of .272 (suggesting low to

187 moderate feature overlap) and an LSA score of .517 (suggesting moderate thematic overlap).

188 As such, some of the overconfidence in JOLs may be attributed more than just item

189 associations. Paired items may also be connected by similar themes or share certain features,

190 resulting in inflated JOLs.

191   JOL tasks can then be manipulated to investigate perceptions of word pair

192 relationships by having participants judge how related they believe the stimuli to be (Maki,

193 2007a, 2007b). The judged values generated from this task can then be compared to the

194 normed databases to create a similar accuracy function or correlation as is created in JOL

195 studies. When presented with the item pair, participants are asked to estimate the number

196 of people out of 100 who would provide the target word when shown only the cue (Maki,

197 2007b), which mimics how the association word norms are created through free association

198 tasks. Maki (2007a) investigated such judgments within the context of associative memory

199 by having participants rate how much associative overlap was shared between items and

200 found that responses greatly overestimated the actual overlap strength for pairs that were

201 weak associates, while underestimating strong associates; thus replicating the Koriat and

202 Bjork (2005) findings for judgments on associative memory, rather than on learning.

203   The judgment of associative memory function (JAM) is created by plotting the judged

204 values by the word pair's normed associative strength and calculating a fit line, which

205 characteristically has a high intercept (bias) with a shallow slope (sensitivity). Figure 1

206 illustrates this function. Overall, the JAM function has been found to be highly reliable and

207 generalized well across multiple variations of the study, with item characteristics such as

208 word frequency, cue set size (QSS), and semantic similarity all having a minimal influence on

209 it (Maki, 2007b). Furthermore, an applied meta-analysis of more than ten studies on JAM

210 indicated that bias and sensitivity are nearly unchangeable, often hovering around 40-60

211 points for the intercept and .20-.30 for the slope (Valentine & Buchanan, 2013). Additionally,

212 Valentine and Buchanan (2013) extended this research to include judgments of semantic

213 memory with the same results.

214   The present study combined PAL and JAM to examine item recall within the context

215 of item judgments, while extending the Maki's JAM task to include judgments of both

216 semantic and thematic memory. Relationship strengths between word pairs were

217 manipulated across each of the three types of memory investigated. Previous research on

²¹⁸ normed databases was used to assure a range of item relatedness. We tested the following

²¹⁹ hypotheses:

1) First, we sought to expand previous Maki (2007b), Maki (2007a), Buchanan (2010), and Valentine and Buchanan (2013) research to include three types of judgments of memory in one experiment, while replicating JAM bias and sensitivity findings. We used the three database norms for association, semantics, and thematics to predict each type of judgment and calculated average slope and intercept values for each participant. We expected to find slope and intercept values that were significantly different from zero, as well as within the range of previous findings. Additionally, we examined the frequency of each predictor being the strongest variable to predict its own judgment condition (i.e., how often association was the strongest predictor of associative judgments, etc.).

2) Given the overlap in these variables and predictions from bidirectional models, we expected to find an interaction between database norms in predicting participant judgments, controlling for judgment type. We used multilevel modeling to examine that interaction of database norms for association, semantics, and thematics in relation to participant judgments.

3) These analyses were then extended to recall as the dependent variable of interest. We examined the interaction of database norms in predicting recall by using a multilevel logistic regression, while controlling for judgment type and rating. We expected to find that database norms would show differences in recall based on the levels other variables (the interaction would be significant), and that ratings would also positively predict recall (i.e., words that participants thought were more related would be remembered better).

4) Finally, we examined if the judgment slopes from Hypothesis 1 would be predictive of recall. Hypothesis 3 examined the direct relationship of word relatedness on recall,

₂₄₄   while this hypothesis explored if participant sensitivity to word relatedness was a

₂₄₅   predictor of recall. For this analysis, we used a multilevel logistic regression to control

₂₄₆   for multiple judgment slope conditions.

₂₄₇   **Methods**

Participants

₂₄₉   A power analysis was conducted using the *simR* package in *R* (Green & MacLeod,

₂₅₀   2016). This package uses simulations to generate power estimates for mixed linear models

₂₅₁   created from the *lme4* package in *R* (Bates, Mächler, Bolker, & Walker, 2015). The results

₂₅₂   of this analyses suggested a minimum of 35 participants would be required to detect an

₂₅₃   effect. However, because power often tends to be underestimated, we extended participant

₂₅₄   recruitment as funding permitted. In total, 112 participants took part in this study.

₂₅₅   Participants were recruited from Amazon's Mechanical Turk, which is a website that allows

₂₅₆   individuals to host projects and connects them with a large pool of respondents who

₂₅₇   complete them for small amounts of money (Buhrmester, Kwang, & Gosling, 2011).

₂₅₈   Participant responses were screened for a basic understanding of the study's instructions.

₂₅₉   Common reasons for rejecting responses included participants entering related words when

₂₆₀   numerical judgment responses were required, and participants responding to the cue words

₂₆₁   during the recall phase with sentences or phrases instead of individual words. Those that

₂₆₂   completed the study correctly were compensated $1.00 for their participation.

Materials

₂₆₄   The stimuli used were sixty-three words pairs of varying associative, semantic, and

₂₆₅   thematic relatedness which were created from the Buchanan et al. (2013) word norm

₂₆₆   database and website. Associative relatedness was measured with Forward Strength (FSG),

₂₆₇   which is the probability that a cue word will elicit a desired target word (Nelson et al., 2004).

₂₆₈   This variable ranges from zero to one wherein zero indicates no association, while one

indicates that participants would always give a target word in response to the cue word. Semantic relatedness was measured with Cosine (COS), which is a measure of semantic feature overlap (Buchanan et al., 2013; McRae et al., 2005; Vinson & Vigliocco, 2008). This variable ranges from zero to one where zero indicates no shared semantic features between concepts and higher numbers indicate more shared features between concepts. Thematic relatedness was calculated with Latent Semantic Analysis (LSA), which generates a score based upon the co-occurrences of words within a document (Landauer & Dumais, 1997; Landauer et al., 1998). LSA values also range from zero to one, indicates no co-occurrence at the low end and higher co-occurrence with higher values. These values were chosen to represent these categories based on face validity and previous research on how word pair psycholinguistic variables overlap (Maki & Buchanan, 2008).

Stimuli were varied such that each variable included a range of each variable. See Table 1 for stimuli averages, SD, and ranges. A complete list of stimuli can be found at http://osf.io/y8h7v. The stimuli were arranged into three blocks for each judgment condition described below wherein each block contained 21 word pairs. Due to limitations of the available stimuli, blocks were structured so that each one contained seven word pairs of low (0-.33), medium (.34-.66), and high (.67-1.00) COS relatedness. Because of this selection process, FSG and LSA strengths are contingent upon the selected stimuli's COS strengths. We selected stimuli within the cosine groupings to cover a range of FSG and LSA values, but certain combinations are often difficult to achieve. For example, there are only four word-pairs that are both high COS and high FSG, thus limiting the ability to manipulate LSA. The study was built online using Qualtrics, and three surveys were created to counter-balance the order in which blocks appeared. Each word pair appeared counter-balanced across each judgment condition, and stimuli were randomized within each block.

Procedure

The present study was divided into three phases. In the first section, participants were presented with word pairs and were asked to make judgments of how related they believed the words in each pair to be. This judgment phase consisted of three blocks of 21 word pairs which corresponded to one of three types of word pair relationships: associative, semantic, or thematic. Each block was preceded by a set of instructions explaining one of the three types of relationships, and participants were provided with examples which illustrated the type of relationship to be judged. Participants were then presented with the word pairs to be judged. The associative block began by explaining associative memory and the role of free association tasks. Participants were provided with examples of both strong and weak associates. For example, *lost* and *found* and were presented as an example of a strongly associated pair, while *article* was paired with *newspaper*, *the*, and *clothing* to illustrate that words can have many weak associates. The semantic judgment block provided participants with a brief overview of how words are related by meaning and showed examples of concepts with both high and low feature overlap. *Tortoise* and *turtle* were provided as an example of two concepts with significant overlap. Other examples were then provided to illustrate concepts with little or no overlap. For the thematic judgments, participants were provided with an explanation of thematic relatedness. *Tree* is explained to be related to *leaf*, *fruit*, and *branch*, but not *computer*. Participants were then given three concepts (*lost*, *old*, *article*) and were asked to come up with words that they feel are thematically related.

After viewing the examples at the start of the block, participants completed the judgment task. Judgment instructions for each block were contingent upon the type of judgment being elicited. For example, instructions in the associative block asked participants to estimate how many college students out of 100 would respond to the cue word with given target, while instructions for semantic judgments asked participants to indicate the percent of features shared between two concepts. The complete experiment can be found at http://osf.io/y8h7v, which contains the exact instructions given to participants for each

321 block and displays the structure of the study. All judgment instructions were modeled after

322 Buchanan (2010) and Valentine and Buchanan (2013).

323   Participants then rated the relatedness of the word pairs based on the set of

324 instructions that they received. In accordance with previous work on JOLs and JAM, item

325 judgments were made using a scale of zero to one hundred, with zero indicating no

326 relationship, and one hundred indicating a perfect relationship. Participants typed their

327 responses into the survey. Once completed, participants then completed the remaining

328 judgment blocks in the same manner. Each subsequent judgment block changed the type of

329 judgment being made. Three versions of the study were created, which counter-balanced the

330 order in which the judgment blocks appeared, and participants were randomly assigned to a

331 survey version. This resulted in each word pair receiving judgments on each of the three

332 types relationships. After completing this section, participants were then presented with a

333 short distractor task to account for recency effects. In this section, participants were

334 presented with a randomized list of the fifty U.S. states and were asked to arrange them in

335 alphabetical order. This task was timed to last two minutes. Once time had elapsed,

336 participants automatically progressed to the final section, which consisted of a cued-recall

337 task. Participants were presented with each of the 63 cue words from the judgment section

338 and were asked to complete each word pair by responding with the correct target word.

339 Participants were informed that they would not be penalized for guessing. The cued-recall

340 task included all stimuli in a random order.

Results

342 **Data Processing and Descriptive Statistics**

343   First, the recall portion of the study was coded as zero for incorrect responses, one for

344 correct responses, and NA for participants who did not complete the recall section (all or

345 nearly all responses were blank). All word responses to judgment items were deleted and set

346 to missing data. The final dataset was created by splitting the initial data file into six

347 sections (one for each of the three experimental blocks and their corresponding recall scores).

348 Each section was individually melted using the *reshape* package in *R* (Wickham, 2007) and

349 was written as a csv file. The six output files were then combined to form the final dataset.

350 Code is available on our OSF page embedded inline with the manuscript in an *R* markdown

351 document (Aust & Barth, 2017). With 112 participants, the dataset in long format included

352 7,056 rows of potential data (i.e., 112 participants * 63 judgments). One incorrect judgment

353 data point (> 100) was corrected to NA. Missing data for judgments or recall were then

354 excluded from the analysis, which includes word responses to judgment items (i.e.,

355 responding with *cat* instead of a number). These items usually excluded a participant from

356 receiving Amazon Mechanical Turk payment, but were included in the datasets found online.

357 In total, 787 data points were excluded (188 judgment only, 279 recall only, 320 both),

358 leading to a final *N* of 105 participants and 6,269 observations. Recall and judgment scores

359 were then screened for outliers using Mahalanobis distance at $p < .001$, and no outliers were

360 found (Tabachnick & Fidell, 2007). To screen for multicollinearity, we examined correlations

361 between judgment items, COS, LSA, and FSG. All correlations were rs $< .50$.

362   The mean judgment of memory for the associative condition ($M = 58.74$, $SD = 30.28$)

363 was lower than the semantic ($M = 66.98$, $SD = 28.31$) and thematic ($M = 71.96$, $SD = $

364 27.80) judgment conditions. Recall averaged over 60% for all three conditions: associative *M*

365 $= 63.40$, $SD = 48.18$; semantic $M = 68.02$, $SD = 46.65$; thematic $M = 64.89$, $SD = 47.74$.

Hypothesis 1

367   Our first hypothesis sought to replicate bias and sensitivity findings from previous

368 research while expanding the JAM function to include judgments based on three types of

369 memory. FSG, COS, and LSA were used to predict each type of judgment. Judgment values

370 were divided by 100, so as to place them on the same scale as the database norms. Slopes

371 and intercepts were then calculated for each participant's ratings for each of the three

372 judgment conditions, as long as they contained at least nine data points out of the 21 that

373 were possible. Single sample *t*-tests were then conducted to test if slope and intercept values

374 significantly differed from zero. See Table 2 for means and standard deviations. Slopes were

375 then compared to the JAM function, which is characterized by high intercepts (between 40

376 and 60 on a 100 point scale) and shallow slopes (between 20 and 40). Because of the scaling

377 of our data, to replicate this function, we should expect to find intercepts ranging from .40 to

378 .60 and slopes in the range of 0.20. to 0.40. Intercepts for associative, semantic, and

379 thematic judgments were each significant, and all fell within or near the expected range.

380 Thematic judgments had the highest intercept at .656, while associative judgments had the

381 lowest intercept at .511.

382   The JAM slope was successfully replicated for FSG in the associative judgment

383 condition, with FSG significantly predicting association, although the slope was slightly

384 higher than expected at .491. COS and LSA did not significantly predict association. For

385 semantic judgments, each of the three database norms were significant predictors. However,

386 JAM slopes were not replicated for this judgment type, as FSG had the highest slope at .118,

387 followed by LSA .085, and then COS .059. These findings were mirrored for thematic

388 judgments, as each database norm was a significant predictor, yet slopes for each predictor

389 fell below range of the expected JAM slopes. Again, FSG had the highest slope, this time

390 just out of range at .192, followed closely by LSA at .188. Interestingly, COS slopes were

391 found to be negative for this judgment condition, -.081. Overall, although JAM slopes were

392 not successfully replicated in each judgment type, the high intercepts and shallow slopes

393 present in all three judgment conditions are still indicative of overconfidence and insensitivity

394 in participant judgments.

395   Additionally, we examined the frequency that each predictor was the maximum

396 strength for each judgment condition. For the associative condition, FSG was the strongest

397 predictor for 64.0 of the participants, with COS and LSA being the strongest for only 16.0

398 and 20.0 of participants respectively. These differences were less distinct when examining the

399 semantic and thematic judgment conditions. In the semantic condition, FSG was highest at

44.1 of participants, LSA was second at 32.4, and COS was least likely at 23.5. Finally, in

the thematic condition, LSA was most likely to be the strongest predictor with 44.6 of

participants, with FSG being the second most likely at 36.6, and COS again being least

likely at 18.8. Interestingly, in all three conditions, COS was least likely to be the strongest

predictor, even in the semantic judgment condition.

Hypothesis 2

As a result of the overlap between variables in Hypothesis 1, the goal of Hypothesis 2

was to test for an interaction between the three database norms when predicting participant

judgment ratings. First, the database norms were mean centered to control for

multicollinearity. The *nlme* package and *lme* function were used to calculate these analyses

(Pinheiro, Bates, Debroy, Sarkar, & R Core Team, 2017). A maximum likelihood multilevel

model was used to test the interaction between FSG, COS, and LSA when predicting

judgment ratings while controlling for type of judgment, with participant number being used

as the random intercept factor. Multilevel models were used to retain all data points (rather

than averaging over items and conditions), while controlling for correlated error due to

participants, as these models are advantageous for multiway repeated measures designs

(Gelman, 2006). This analysis resulted in a significant three-way interaction between FSG,

COS, and LSA ($\beta = 3.324$, $p < .001$), which is examined below in a simple slopes analysis.

Table 3 includes values for main effects, two-way, and three-way interactions.

To investigate this interaction, simple slopes were calculated for low, average, and high

levels of COS. This variable was chosen for two reasons: first, it was found to be the weakest

of the three predictors in hypothesis one, and second, manipulating COS would allow us to

track changes across FSG and LSA. Significant two-way interactions were found between

FSG and LSA at both low COS ($\beta = -1.492$, $p < .001$), average COS ($\beta = -0.569$, $p < .001$),

and high COS ($\beta = 0.355$, $p = .013$). A second level was then added to the analysis in which

simple slopes were created for each level of LSA, allowing us to assess the effects of LSA at

different levels of COS on FSG. When both COS and LSA were low, FSG significantly

predicted judgment ratings ($\beta = 0.663$, p < .001). At low COS and average LSA, FSG

decreased but still significantly predicted judgment ratings ($\beta = 0.375$, p < .001). However,

when COS was low and LSA was high, FSG was not a significant predictor ($\beta = 0.087$, p =

.079). A similar set of results was found at the average COS level. When COS was average

and LSA was LOW, FSG was a significant predictor, ($\beta = 0.381$, $p < .001$). As LSA

increased at average COS levels, FSG decreased in strength: average COS, average LSA FSG

($\beta = 0.355$, p .013) and average COS, high LSA FSG ($\beta = 0.161$, $p < .001$). This finding

suggests that at low COS, LSA and FSG create a seesaw effect in which increasing levels of

thematics is counterbalanced by decreasing importance of association when predicting

judgments. FSG was not a significant predictor when COS was high and LSA was low (

0.099, p = .088). At high COS and average LSA, FSG significantly predicted judgment

ratings ($\beta = 0.167$, p < .001), and finally when both COS and LSA were high, FSG

increased and was a significant predictor of judgment ratings ($\beta = 0.236$, p < .001). Thus, at

high levels of COS, FSG and LSA are complementary when predicting recall, increasing

together as COS increases. Figure 2 displays the three-way interaction wherein the top row

of figures indicates the seesaw effect, as LSA increases FSG decreases in strength. The

bottom row indicates the complementary effect where increases in LSA occur with increases

in FSG predictor strength.

Hypothesis 3

Given the results of Hypothesis 2, we then sought to extend the analysis to participant

recall scores. A multilevel logistic regression was used with the *lme4* package and *glmer()*

function (Pinheiro et al., 2017), testing the interaction between FSG, COS, and LSA when

predicting participant recall. As with the previous hypothesis, we controlled for type of

judgement and, additionally, covaried judgment ratings. Participants were used as a random

intercept factor. Judged values were a significant predictor of recall, ($\beta = 0.686$, $p < .001$)

452 where increases in judged strength predicted increases in recall. A significant three-way

453 interaction was detected between FSG, COS, and LSA ($\beta = 24.572$, $p < .001$). See Table 4

454 for main effects, two-way, and three-way interaction values.

455   The moderation process from Hypothesis 2 was then repeated, with simple slopes first

456 calculated at low, average, and high levels of COS. This set of analyses resulted in significant

457 two-way interactions between LSA and FSG at low COS ($\beta = -7.845$, $p < .001$) and high

458 COS ($\beta = 5.811$, $p = .009$). No significant two-way interaction was found at average COS (β

459 $= -1.017$, $p = .493$). Following the design of hypothesis two, simple slopes were then

460 calculated for low, average, and high levels of LSA at the low and high levels of COS,

461 allowing us to assess how FSG effects recall at varying levels of both COS and LSA. When

462 both COS and LSA were low, FSG was a significant predictor of recall ($\beta = 4.116$, $p < .001$).

463 At low COS and average LSA, FSG decreased from both low levels, but was still a significant

464 predictor ($\beta = 2.601$, $p < .001$), and finally, low COS and high LSA, FSG was the weakest

465 predictor of the three ($\beta = 1.086$, $p = .030$). As with Hypothesis 2, LSA and FSG

466 counterbalanced one another, wherein the increasing levels of thematics led to a decrease in

467 the importance of association in predicting recall. At high COS and low LSA, FSG was a

468 significant predictor ($\beta = 2.447$, $p = .003$). When COS was high and LSA was average, FSG

469 increased as a predictor and remained significant ($\beta = 3.569$, $p < .001$). This finding

470 repeated when both COS and LSA were high, with FSG increasing as a predictor of recall (β

471 $= 4.692$, $p < .001$). Therefore, at high levels of COS, LSA and FSG are complementary

472 predictors of recall, increasing together and extending the findings of Hypothesis 2 to

473 participant recall. Figure 3 displays the three-way interaction. The top left figure indicates

474 the counterbalancing effect of recall of LSA and FSG, while the top right figure shows no

475 differences in simple slopes for average levels of cosine. The bottom left figure indicates the

476 complementary effects where LSA and FSG increase together as predictors of recall at high

477 COS levels.

Hypothesis 4

In our fourth and final hypothesis, we investigated whether the judgment slopes and intercepts obtained in Hypothesis 1 would be predictive of recall ability. Whereas Hypothesis 3 indicated that word relatedness was directly related to recall performance, this hypothesis instead looked at whether or not participants' sensitivity and bias to word relatedness could be used a predictor of recall (Maki, 2007b). This analysis was conducted with a multilevel logistic regression, as described in Hypothesis 3 where each database slope and intercept was used as predictors of recall using participant as a random intercept factor. These analyses were separated by judgment type, so that each set of judgment slopes and intercepts were used to predict recall. The separation controlled for the number of variables in the equation, as all slopes and intercepts would have resulted in overfitting. These values were obtained from Hypothesis 1 where each participant's individual slopes and intercepts were calculated for associative, semantic, and thematic judgment conditions. Table 2 shows average slopes and intercepts for recall for each of the three types of memory, and Table 5 portrays the regression coefficients and statistics. In the associative condition, FSG slope significantly predicted recall ($b = 0.898$, $p = .008$), while COS slope ($b = 0.314$, $p = .568$) and LSA slope ($b = 0.501$, $p = .279$) were non-significant. In the semantic condition, COS slope ($b = 2.039$, $p < .001$) and LSA slope ($b = 1.061$, $p = .020$) were both found to be significant predictors of recall. FSG slope was non-significant in this condition ($b = 0.381$, $p = .187$). Finally, no predictors were significant in the thematic condition, though LSA slope was found to be the strongest ($b = 0.896$, $p = .090$).

Discussion

This study investigated the relationship between associative, semantic, and thematic word relations and their effect on participant judgments and recall performance through the testing of four hypotheses. In our first hypothesis, bias and sensitivity findings first proposed by Maki (2007a) were successfully replicated in the associative condition, with slope and

504 intercept values falling within the expected range. While these findings were not fully

505 replicated when extending the analysis to include semantic and thematic judgments (as

506 slopes in these conditions did not fall within the appropriate range), participants still

507 displayed high intercepts and shallow slopes, suggesting overconfidence in judgment making

508 and an insensitivity to changes in strength between pairs. Additionally, when looking at the

509 frequency that each predictor was the strongest in making these judgments, FSG was the

510 best predictor for both the associative and semantic conditions, while LSA was the best

511 predictor in the thematic condition. In each of the three conditions, COS was the weakest

512 predictor, even when participants were asked to make semantic judgments. This finding

513 suggests that associative relationships seem to take precedence over semantic relationships

514 when judging pair relatedness, regardless of what type of judgment is elicited. Additionally,

515 this finding may be taken as further evidence of a separation between associative information

516 and semantic information, in which associative information is always processed, while

517 semantic information may be suppressed due to task demands (Buchanan, 2010; Hutchison

518 & Bosco, 2007).

519   Our second hypothesis examined the three-way interaction between FSG, COS, and

520 LSA when predicting participant judgments. At low semantic overlap, a seesaw effect was

521 found in which increases in thematic strength led to decreases in associative predictiveness.

522 This finding was then replicated in hypothesis 3 when extending the analysis to predict

523 recall. By limiting the semantic relationships between pairs, an increased importance is

524 placed on the role of associations and thematics when making judgments or retrieving pairs.

525 In such cases, increasing the amount of thematic overlap between pairs results in thematic

526 relationships taking precedent over associative relationships. However, when semantic

527 overlap was high, a complementary relationship was found in which increases in thematic

528 strength in turn led to increases in the strength of FSG as a predictor. This result suggests

529 that at high semantic overlap, associations and thematic relations build upon one another.

530 Because thematics is tied to both semantic overlap and item associations, the presence of

531 strong thematic relationships between pairs during conditions of high semantic overlap

532 boosts the predictive ability of associative word norms. Again, this complementary effect was

533 found when examining both recall and judgments.

534   Finally, our fourth hypothesis used the judgment slopes and intercepts calculated in

535 Hypothesis 1 to investigate if participants' bias and sensitivity to word relatedness could be

536 used to predict recall. For the associative condition, the FSG slope significantly predicted

537 recall. In the semantic condition, recall was significantly predicted by both the COS and

538 LSA slopes. However, for the thematic condition, although the LSA slope was the strongest,

539 no predictors were significant. One explanation for this finding is that thematic relationships

540 between item pairs act as a blend between associations and semantics. As such, LSA faces

541 increased competition from the associative and semantic database norms when predicting

542 recall in this manner.

543   Overall, our findings indicated the degree to which the processing of associative,

544 semantic, and thematic information impacts retrieval and judgment making tasks and the

545 interactive relationship that exists between these three types of lexical information. While

546 previous research has shown that memory networks are divided into separate systems which

547 handle storage and processing for meaning and association, the presence of these interactions

548 suggests that connections exist between these networks, linking them to one another. As

549 such, we suggest that these memory systems may form a three-tiered, interconnected system.

550 First, information enters the semantic memory network, which processes features of concepts

551 and provides a means of categorizing items based on the similarity of their features. Next,

552 the associative network adds information for items based on contexts generated by reading or

553 speech. Finally, the thematic network pulls in information from both the semantic and

554 associative networks to create a mental representation of both the item and its place world

555 relative to other concepts. This study did not explore the timing of information input from

556 each of these systems, but it may be similar to a dual-route model of reading and naming, in

557 that each runs in parallel contributing the judgment and recall process (Coltheart, Curtis,

Atkins, & Haller, 1993).

Viewing this model purely through the lens of semantic memory, it draws comparison to dynamic attractor models (Hopfield, 1982; Jones et al., 2015; McLeod et al., 2000). One of the defining features of dynamic attractor models is that they allow for some type of bidirectionally or feedback between connections in the network. In the study of semantic memory, these models are useful for taking into account multiple restraints such as links between semantics and the orthography of the concept in question. Our hypothesis extends this notion as a means of framing how these three memory systems are connected. The underlying meaning of a concept is linked with both information pertaining to its co-occurrences in everyday language and information relating to the general contexts in which it typically appears.

How then does this hypothesis lend itself towards the broader context of psycholinguistic research? One application of this hypothesis may be models of word recognition. One popular model is Seidenberg and Mcclelland (1989) "triangle model", and several variations of this model have been proposed and tested (see Harley, 2008 for a review). This model recognizes speech and reading based upon the orthography, phonology, and meaning of words. Each of these three word properties are linked to in such a way that orthography is linked to phonology, phonology is linked with meaning, and meaning is linked to orthography (forming a triangle). The pathways between word properties are bidirectional, allowing for feedback between connections. Whereas the original version of this model focused almost exclusively on the link between orthography and phonology, Harm and Seidenberg (2004) developed a version which included a focus on semantics, with word meaning being based on input from the orthography and phonology components of the model. Future studies in this area may wish to incorporate thematic and associative knowledge as elements of meaning, as thematic and associative information is interconnected with the semantic network. Ultimately, further studies will be needed to explore the interconnections between the semantic, thematic, and associative networks.

References

Aust, F., & Barth, M. (2017). *papaja: Create APA manuscripts with R Markdown.*
     Retrieved from https://github.com/crsh/papaja

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models
     using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

Buchanan, E. M. (2010). Access into memory: Differences in judgments and priming for
     semantic and associative memory. *Journal of Scientific Psychology*, *March*, 1–8.

Buchanan, E. M., Holmes, J. L., Teasley, M. L., & Hutchison, K. A. (2013). English
     semantic word-pair norms and a searchable Web portal for experimental stimulus
     creation. *Behavior Research Methods*, *45*(3), 746–757. doi:10.3758/s13428-012-0284-z

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk.
     Perspectives on Psychological Science, *6*(1), 3–5. doi:10.1177/1745691610393980

Chow, B. W.-Y. (2014). The differential roles of paired associate learning in Chinese and
     English word reading abilities in bilingual children. *Reading and Writing*, *27*(9),
     1657–1672. doi:10.1007/s11145-014-9514-3

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud:
     Dual-route and parallel-distributed-processing approaches. *Psychological Review*,
     100(4), 589–608. doi:10.1037//0033-295X.100.4.589

Gelman, A. (2006). Multilevel (hierarchical) modeling: What it can and cannot do.
     Technometrics, *48*(3), 432–435. doi:10.1198/004017005000000661

Green, P., & MacLeod, C. J. (2016). SIMR: an R package for power analysis of generalized
     linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498.
     doi:10.1111/2041-210X.12504

Harley, T. (2008). *The psychology of language: From data to theory* (3rd ed.). New York:
     Psychology Press.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading:
     Cooperative division of labor between visual and phonological processes.

Psychological Review, 111(3), 662–720. doi:10.1037/0033-295X.111.3.662

Hertzog, C., Kidder, D. P., Powell-Moman, A., & Dunlosky, J. (2002). Aging and monitoring associative learning: Is monitoring accuracy spared or impaired? *Psychology and Aging, 17*(2), 209–225. doi:10.1037/0882-7974.17.2.209

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, 79*(8), 2554–2558. doi:10.1073/pnas.79.8.2554

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review, 10*(4), 785–813. doi:10.3758/BF03196544

Hutchison, K. A., & Bosco, F. A. (2007). Congruency effects in the letter search task: semantic activation in the absence of priming. *Memory & Cognition, 35*(3), 514–525. doi:10.3758/BF03193291

Jiang, J. J., & Conrath, D. W. (1997). Semantic similarity based on corpus statistics and lexical taxonomy. *Proceedings of International Conference Research on Computational Linguistics*, 19–33. Retrieved from http://arxiv.org/abs/cmp-lg/9709008

Jones, L. L., & Golonka, S. (2012). Different influences on lexical priming for integrative, thematic, and taxonomic relations. *Frontiers in Human Neuroscience, 6*, 1–17. doi:10.3389/fnhum.2012.00205

Jones, M. N., Willits, J., & Dennis, S. (2015). Models of semantic memory. In A. T. Townsend & Jerome R. Busemeyer (Eds.), *Oxford handbook of mathematical and computational psychology* (pp. 232–254). Oxford University Press. doi:10.1093/oxfordhb/9780199957996.013.11

Jouravlev, O., & McRae, K. (2016). Thematic relatedness production norms for 100 object concepts. *Behavior Research Methods, 48*(4), 1349–1357. doi:10.3758/s13428-015-0679-8

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge

during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 187–194. doi:10.1037/0278-7393.31.2.187

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. doi:10.1037//0033-295X.104.2.211

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2), 259–284. doi:10.1080/01638539809545028

Lucas, M. (2000). Semantic priming without association: a meta-analytic review. *Psychonomic Bulletin & Review*, *7*(4), 618–630. doi:10.3758/BF03212999

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*(2), 203–208. doi:10.3758/BF03204766

Maki, W. S. (2007a). Judgments of associative memory. *Cognitive Psychology*, *54*(4), 319–353. doi:10.1016/j.cogpsych.2006.08.002

Maki, W. S. (2007b). Separating bias and sensitivity in judgments of associative memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 231–237. doi:10.1037/0278-7393.33.1.231

Maki, W. S., & Buchanan, E. M. (2008). Latent structure in measures of associative, semantic, and thematic knowledge. *Psychonomic Bulletin & Review*, *15*(3), 598–603. doi:10.3758/PBR.15.3.598

Maki, W. S., McKinley, L. N., & Thompson, A. G. (2004). Semantic distance norms computed from an electronic dictionary (WordNet). *Behavior Research Methods, Instruments, & Computers*, *36*(3), 421–431. doi:10.3758/BF03195590

McLeod, P., Shallice, T., & Plaut, D. C. (2000). Attractor dynamics in word recognition: converging evidence from errors by normal subjects, dyslexic patients and a connectionist model. *Cognition*, *74*(1), 91–114. doi:10.1016/S0010-0277(99)00067-0

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature

production norms for a large set of living and nonliving things. *Behavior Research Methods*, *37*(4), 547–559. doi:10.3758/BRM.40.1.183

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227–234. doi:10.1037/h0031564

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1975). Loci of contextual effects on visual word-recognition. In P. M. A. Rabbitt (Ed.), *Attention and performance v.* London, UK: Academic Press.

Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, *38*(11), 39–41. doi:10.1145/219717.219748

Nelson, D. L., McEvoy, C. L., & Dennis, S. (2000). What is free association and what does it measure? *Memory & Cognition*, *28*(6), 887–899. doi:10.3758/BF03209337

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–407. doi:10.3758/BF03195588

Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, *76*(3), 241–263. doi:10.1037/h0027272

Pinheiro, J., Bates, D., Debroy, S., Sarkar, D., & R Core Team. (2017). nlme: Linear and Nonlinear Mixed Effects Models. Retrieved from https://cran.r-project.org/package=nlme

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*(1), 56–115. doi:10.1037//0033-295X.103.1.56

Richardson, J. T. E. (1998). The availability and effectiveness of reported mediators in associative learning: A historical review and an experimental investigation. *Psychonomic Bulletin & Review*, *5*(4), 597–614. doi:10.3758/BF03208837

Riordan, B., & Jones, M. N. (2011). Redundancy in perceptual and linguistic experience:

Comparing feature-based and distributional models of semantic representation. *Topics in Cognitive Science*, *3*(2), 303–345. doi:10.1111/j.1756-8765.2010.01111.x

Rogers, T. T., & McClelland, J. L. (2006). *Semantic cognition.* Cambridge, MA: MIT Press.

Rumelhart, D. E., McClelland, J. L., & Group, P. R. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1.* Cambridge, MA: MIT Press.

Schwartz, B. L., & Brothers, B. R. (2013). Survival processing does not improve paired-associate learning. In B. L. Schwartz, M. L. Howe, M. P. Toglia, & H. Otgaar (Eds.), *What is adaptive about adaptive memory?* (pp. 159–171). Oxford University Press. doi:10.1093/acprof:oso/9780199928057.003.0009

Seidenberg, M. S., & Mcclelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*, 523–568.

Smythe, P. C., & Paivio, A. (1968). A comparison of the effectiveness of word imagery and meaningfulness in paired-associate learning of nouns. *Psychonomic Science*, *10*(2), 49–50. doi:10.3758/BF03331401

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics* (5th ed.). New York, NY: Allyn & Bacon.

Valentine, K. D., & Buchanan, E. M. (2013). JAM-boree: An application of observation oriented modelling to judgements of associative memory. *Journal of Cognitive Psychology*, *25*(4), 400–422. doi:10.1080/20445911.2013.775120

Vinson, D. P., & Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behavior Research Methods*, *40*(1), 183–190. doi:10.3758/BRM.40.1.183

Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, *21*(12). doi:10.18637/jss.v021.i12

Table 1

Summary Statistics for Stimuli

Variable	COS Low			COS Average			COS High		
	N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
COS	21	.115	.122	21	.461	.098	21	.754	.059
FSG Low	18	.062	.059	18	.122	.079	17	.065	.067
FSG Average	3	.413	.093	2	.411	.046	2	.505	.175
FSG High	NA	NA	NA	1	.697	NA	2	.744	.002
LSA Low	16	.174	.090	8	.220	.074	7	.282	.064
LSA Average	5	.487	.126	10	.450	.111	12	.478	.095
LSA High	NA	NA	NA	3	.707	.023	2	.830	.102

Note. COS: Cosine, FSG: Forward Strength, LSA: Latent Semantic Analysis.

Table 2

Summary Statistics for Hypothesis 1 t-Tests

Variable	M	SD	t	df	p	d	95%CI
Associative Intercept	.511	.245	20.864	99	< .001	2.086	1.734 - 2.435
Associative COS	-.030	.284	-1.071	99	.287	-0.107	-0.303 - 0.090
Associative FSG	.491	.379	12.946	99	< .001	1.295	1.027 - 1.559
Associative LSA	.035	.317	1.109	99	.270	0.111	-0.086 - 0.307
Semantic Intercept	.587	.188	31.530	101	< .001	3.122	2.649 - 3.592
Semantic COS	.059	.243	2.459	101	.016	0.244	0.046 - 0.440
Semantic FSG	.118	.382	3.128	101	.002	0.310	0.110 - 0.508
Semantic LSA	.085	.304	2.816	101	.006	0.279	0.080 - 0.476
Thematic Intercept	.656	.186	35.475	100	< .001	3.530	3.002 - 4.048
Thematic COS	-.081	.239	-3.405	100	< .001	-0.339	-0.539 - -0.137
Thematic FSG	.192	.306	6.290	100	< .001	0.626	0.411 - 0.838
Thematic LSA	.188	.265	7.111	100	< .001	0.708	0.488 - 0.924

Note. Confidence interval for d was calculated using the non-central t-distribution.

Table 3

MLM Statistics for Hypothesis 2

Variable	beta	SE	t	p
Intercept	0.603	0.014	43.287	< .001
Semantic Judgments	0.079	0.008	9.968	< .001
Thematic Judgments	0.127	0.008	16.184	< .001
ZCOS	-0.103	0.017	-6.081	< .001
ZLSA	0.090	0.022	4.196	< .001
ZFSG	0.271	0.029	9.420	< .001
ZCOS:ZLSA	-0.141	0.085	-1.650	.099
ZCOS:ZFSG	-0.374	0.111	-3.364	< .001
ZLSA:ZFSG	-0.569	0.131	-4.336	< .001
ZCOS:ZLSA:ZFSG	3.324	0.490	6.791	< .001
Low COS ZLSA	0.129	0.033	3.934	< .001
Low COS ZFSG	0.375	0.049	7.679	< .001
Low COS ZLSA:ZFSG	-1.492	0.226	-6.611	< .001
High COS ZLSA	0.051	0.031	1.647	.100
High COS ZFSG	0.167	0.034	4.878	< .001
High COS ZLSA:ZFSG	0.355	0.143	2.484	.013
Low COS Low LSA ZFSG	0.663	0.078	8.476	< .001
Low COS High LSA ZFSG	0.087	0.049	1.754	.079
Avg COS Low LSA ZFSG	0.381	0.047	8.099	< .001
Avg COS High LSA ZFSG	0.161	0.027	5.984	< .001
High COS Low LSA ZFSG	0.099	0.058	1.707	.088
High COS High LSA ZFSG	0.236	0.023	10.263	< .001

Note. Database norms were mean centered. The table shows
main effects and interactions for database norms at low, average,
and high levels of COS and LSA when predicting participant
judgments.

Table 4

MLM Statistics for Hypothesis 3

Variable	beta	SE	z	p
Intercept	0.301	0.138	2.188	.029
Semantic Judgments	0.201	0.074	2.702	.007
Thematic Judgments	-0.001	0.075	-0.020	.984
Judged Values	0.686	0.115	5.956	< .001
ZCOS	0.594	0.179	3.320	< .001
ZLSA	-0.350	0.204	-1.714	.087
ZFSG	3.085	0.302	10.205	< .001
ZCOS:ZLSA	2.098	0.837	2.506	.012
ZCOS:ZFSG	1.742	1.306	1.334	.182
ZLSA:ZFSG	-1.017	1.484	-0.685	.493
ZCOS:ZLSA:ZFSG	24.572	6.048	4.063	< .001
Low COS ZLSA	-0.933	0.301	-3.099	.002
Low COS ZFSG	2.601	0.471	5.521	< .001
Low COS ZLSA:ZFSG	-7.845	2.204	-3.560	< .001
High COS ZLSA	0.233	0.317	0.737	.461
High COS ZFSG	3.569	0.470	7.586	< .001
High COS ZLSA:ZFSG	5.811	2.231	2.605	.009
Low COS Low LSA ZFSG	4.116	0.741	5.558	< .001
Low COS High LSA ZFSG	1.086	0.501	2.166	.030
High COS Low LSA ZFSG	2.447	0.811	3.018	.003
High COS High LSA ZFSG	4.692	0.388	12.083	< .001

Note. Database norms were mean centered. The table shows
main effects and interactions for database norms at low, average,
and high levels of COS and LSA when predicting recall.

Table 5

MLM Statistics for Hypothesis 4

Variable	b	SE	z	p
(Intercept)	-0.432	0.439	-0.983	.326
ACOS	0.314	0.550	0.572	.568
ALSA	0.501	0.463	1.081	.279
AFSG	0.898	0.337	2.667	.008
AIntercept	1.514	0.604	2.507	.012
(Intercept)	-0.827	0.463	-1.787	.074
SCOS	2.039	0.518	3.939	< .001
SLSA	1.061	0.455	2.335	.020
SFSG	0.381	0.289	1.319	.187
SIntercept	2.292	0.681	3.363	< .001
(Intercept)	0.060	0.599	0.101	.920
TCOS	0.792	0.566	1.401	.161
TLSA	0.896	0.529	1.694	.090
TFSG	-0.394	0.441	-0.894	.371
TIntercept	1.028	0.756	1.360	.174

Note. Each judgment-database bias and sensitivity predicting recall for corresponding judgment block. A: Associative, S: Semantic, T: Thematic.
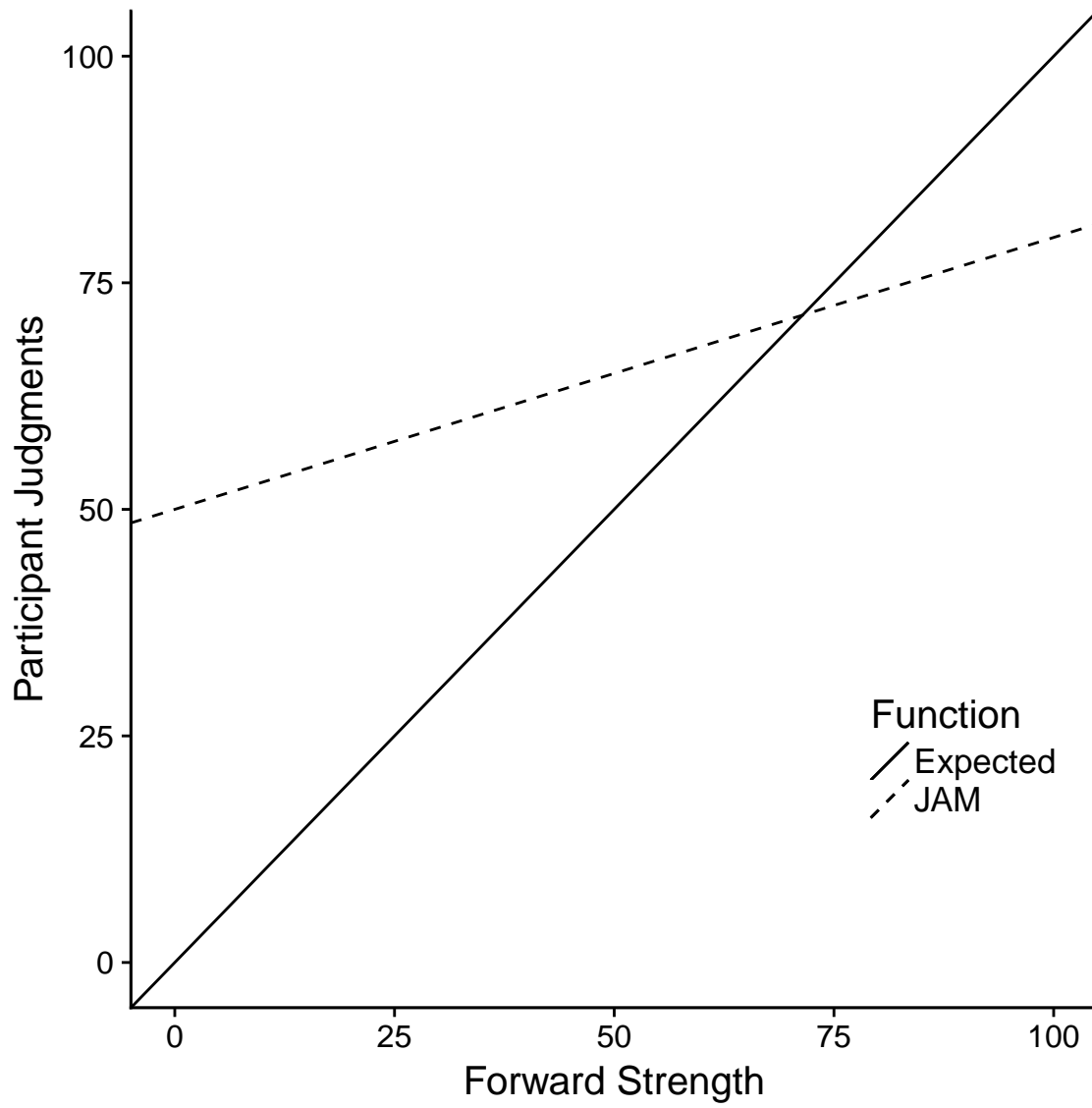
Figure 1. JAM slope findings from Maki (2007a). JAM is characterized by a high intercept (between 40 and 60) and a shallow slope (between 0.20 and 0.40). The solid line shows expected results if judgment ratings are perfectly calibrated with association norms.
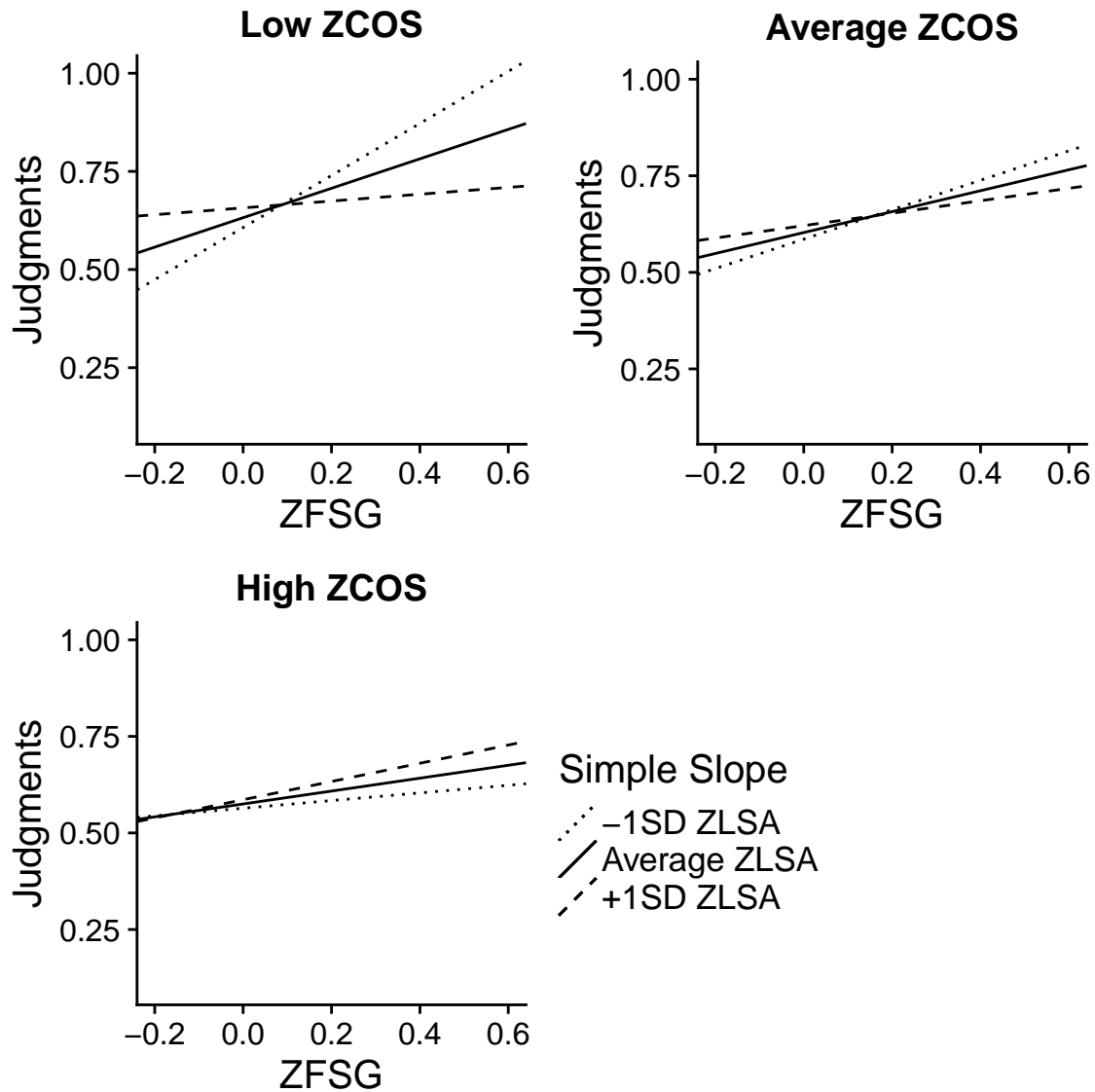
Figure 2. Simple slopes graph displaying the slope of FSG when predicting participant judgments at low, average, and high LSA split by low, average, and high COS. All variables were mean centered.
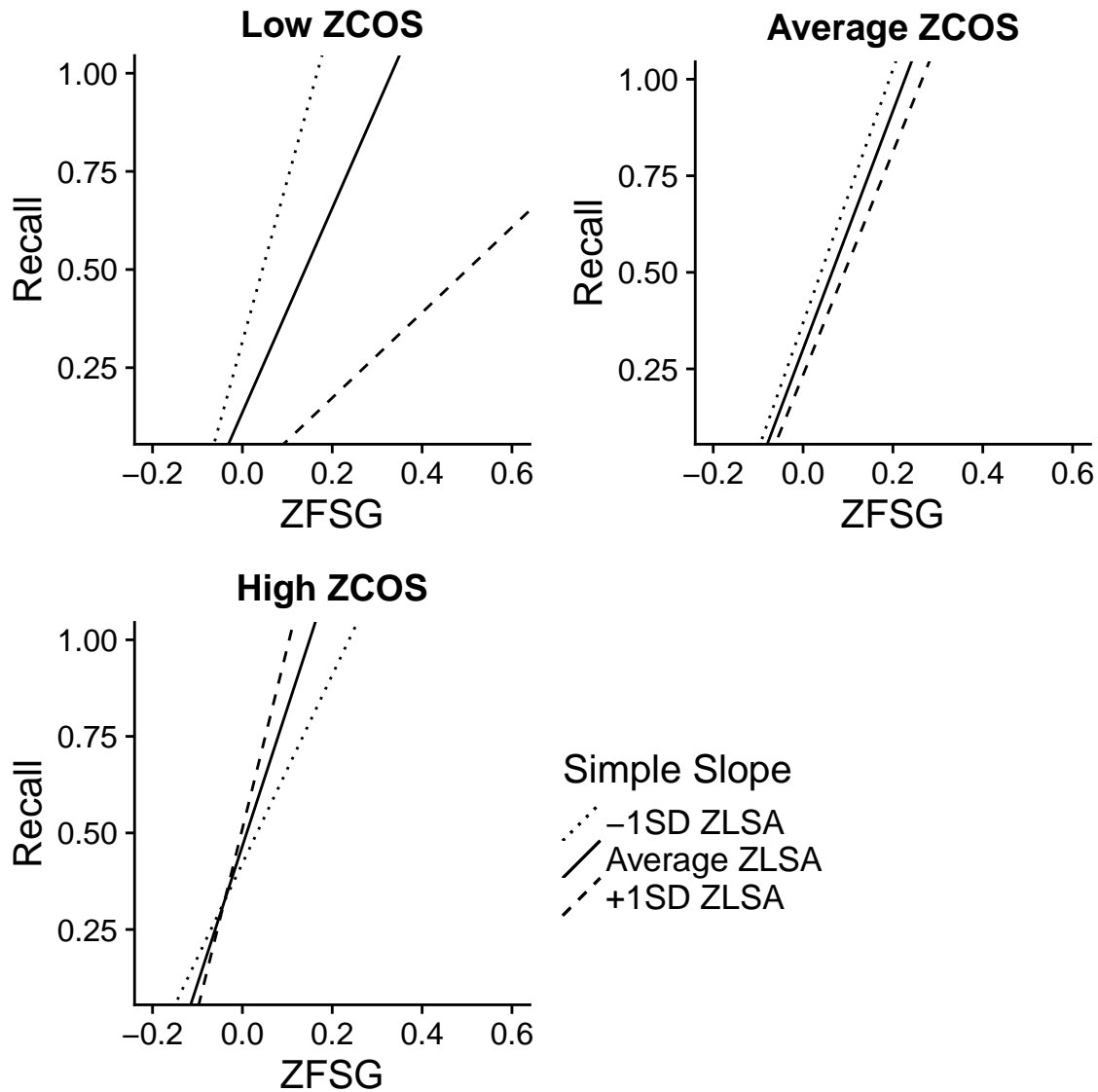
Figure 3. Simple slopes graph displaying the slope of FSG when predicting recall at low, average, and high LSA split by low, average, and high COS. All variables were mean centered.