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

2   for Memory Judgments and Retrieval

3   Nicholas P. Maxwell[1,2] & Erin M. Buchanan[1]

4   [1] Missouri State University

5   [2] University of Southern Mississippi

12   Abstract

13   This study examined the interactive relationship between semantic, thematic, and

14   associative word pair strength in the prediction of item judgments and cued-recall

15   performance. Participants were recruited from Amazon's Mechanical Turk and were given

16   word pairs of varying relatedness to judge for their semantic, thematic, and associative

17   strength. After completing a distractor task, participants then completed a cued recall task.

18   First, we sought to expand previous work on judgments of associative memory (JAM) to

19   include semantic and thematic based judgments, while also replicating bias and sensitivity

20   findings. Next, we tested for an interaction between the three database norms (FSG, COS,

21   and LSA) when predicting participant judgments and also extended previous work to test for

22   interactions between the three database norms when predicting recall. Significant three-way

23   interactions were found between FSG, COS, and LSA when predicting judgments and recall.

24   For low semantic feature overlap, thematic and associative strength were competitive; as

25   thematic strength increased, associative predictiveness decreased. However, this trend

26   reversed for high semantic feature overlap, wherein thematic and associative strength were

27   complementary as both set of simple slopes increased together. Overall, our findings indicate

28   the degree to which the processing of associative, semantic, and thematic information

29   impacts cognitive processes such as retrieval and item judgments, while also examining the

30   underlying, interactive relationship that exists between these three types of information.

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

₃₂ Investigating the Interaction between Associative, Semantic, and Thematic Database Norms

₃₃   for Memory Judgments and Retrieval

₃₄   The study of cognition has a rich history of exploring the role of association in human

₃₅ memory. One key finding is that elements of cognitive processing play a critical role in how

₃₆ well an individual retains learned information. Throughout the mid-20th century, researchers

₃₇ investigated this notion, particularly through the use of paired-associate learning (PAL). In

₃₈ this paradigm, participants are presented with a pair of items and are asked to make

₃₉ connections between them so that the presentation of one item (the cue) will in turn trigger

₄₀ the recall of the other (the target). Early studies of this nature focused primarily on the

₄₁ effects of meaning and imagery on recall performance. For example, Smythe and Paivio

₄₂ (1968) found that noun imagery played a crucial role in PAL performance; subjects were

₄₃ much more likely to remember word-pairs that were low in meaning similarity if imagery

₄₄ between the two was high. Subsequent studies in this area focused on the effects of

₄₅ mediating variables on PAL tasks as well as the effects of imagery and meaningfulness on

₄₆ associative learning (Richardson, 1998), with modern studies shifting their focus towards a

₄₇ broad range of applied topics such as how PAL is affected by aging (Hertzog, Kidder,

₄₈ Powell-Moman, & Dunlosky, 2002), its impacts on second language acquisition (Chow, 2014),

₄₉ and even in evolutionary psychology (Schwartz & Brothers, 2013).

₅₀   Early PAL studies routinely relied on stimuli generated from word lists that focused

₅₁ extensively on measures of word frequency, concreteness, meaningfulness, and imagery

₅₂ (Paivio, 1969). However, the word pairs in these lists were typically created due to their

₅₃ apparent relatedness or frequency of occurrence in text. While lab self-generation appears

₅₄ face valid, one finds that this method of selection lacks a decisive method of defining the

₅₅ underlying relationships between the pairs (Buchanan, 2010), as these variables only capture

₅₆ psycholinguistic measurements of an individual concept (i.e., how concrete is *cat* and word

₅₇ 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.

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, McEvoy, & Schreiber, 2004). More recently, the Small World of Words project (SWOW; De Deyne, Navarro, & Storms, 2013) has sought to capture associations between Dutch words by employing a multiple response technique in contrast to the traditional single response free association task used by Nelson et al. (2004). These norms are now being collected for English words (De Deyne, Navarro, Perfors, Brysbaert, & Storms, 2018).

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 (e.g., Rogers & McClelland, 2006; Rumelhart, McClelland, & PDP Research 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 based on the weighted strength of the corresponding unit connections (Jones, Willits, & Dennis, 2015). On the other hand, distributional models of semantic memory posit that semantic representations are created through the co-occurrences of words together in a body of text and suggest that words with similar meanings will appear together in similar contexts (Riordan & Jones, 2011). Popular distributional models of semantic memory include Latent Semantic Analysis (LSA; Landauer & Dumais, 1997) and the Hyperspace Analogue to Language (HAL; Lund & Burgess, 1996).

Feature production tasks are a common means of producing semantic word norms (Buchanan, Holmes, Teasley, & Hutchison, 2013; McRae, Cree, Seidenberg, & McNorgan, 2005; Vinson & Vigliocco, 2008) In such tasks, participants are shown the name of a concept and are asked to list what they believe the concept's most important features to be (McRae et al., 2005). Several statistical measures have been developed which measure the degree of feature overlap between concepts. Similarity between any two concepts can be measured by representing them as vectors and calculating the cosine value (COS) between them (Maki, McKinley, & Thompson, 2004). Cosine values range from 0 (unrelated) to 1 (perfectly related). For example, the pair *hornet - wasp* has a COS of .88, indicating a high degree of overlap between the two concepts. Feature overlap can also be measured by JCN, which involves calculating the information content value of each concept and the lowest super-ordinate shared by each concept using an online dictionary, such as WordNET (Miller, 1995). The JCN value is then computed by summing together the difference of each concept and its lowest super-ordinate (Jiang & Conrath, 1997; Maki et al., 2004). The advantage to

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

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

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

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

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

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

117 within categories (e.g., *bird - pigeon*), thematic relationships center around the links between

118 concepts and an overarching theme or scenario (e.g., *bird - nest*; Jones & Golonka, 2012).

119 Jouravlev and McRae (2016) provide a list of 100 thematic relatedness production norms,

120 which were generated through a task similar to feature production in which participants were

121 presented with a concept and were asked to list names of other concepts they believed to be

122 related. Distributional models of semantic memory also lend themselves well to the study of

123 thematic word relations. Because these models are text-based and score word pair relations

124 in regard to their overall context within a document, they assess thematic knowledge as well

125 as semantic knowledge. Additionally, text-based models such as LSA are able to account for

126 both the effects of context and similarity of meaning, bridging the gap between associations

127 and semantics (Landauer, Foltz, Laham, Folt, & Laham, 1998).

128   Discussion of these measures then leads to the question of whether each one truly

129 assesses some unique concept or if they simply tap into our overall linguistic knowledge.

130 Taken at face value, word pair associations and semantic word relations appear to be vastly

131 different, yet the line between semantics/associations and thematics is much more blurred.

132 While thematic word relations are indeed an aspect of semantic memory and include word

133 co-occurrence as an integral part of their creation, themes also appear to be indicative of a

134 separate area of linguistic processing. Previous research by Maki and Buchanan (2008)

135 appears to confirm this theory. Using clustering and factor analysis techniques, they

136 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. For example, some of the first studies on semantic priming used association word norms for stimuli creation (Lucas, 2000; Meyer & Schvaneveldt, 1971; Meyer, Schvaneveldt, & Ruddy, 1975). This observation becomes strikingly apparent when one desires the creation of word pairs related on only one dimension. One particular difficulty faced by researchers comes when attempting to separate association strength from feature overlap, as highly associated items tend to be semantically related as well. Additionally, a lack of association strength between two items may not necessarily be indicative of a total lack of association, as traditional norming tasks typically do not produce a large enough set of responses to capture all available associations between items. Some items with extremely weak associations may inevitably slip through the cracks (Hutchison, 2003). As such, the present study seeks to provide further insight by examining how different levels of associative overlap (measured in FSG), semantic overlap (feature overlap measured with COS), and thematic overlap (measured with LSA) affect cognitive tasks such as short term item retrieval and item relatedness judgments. Instead of focusing solely on one variable or trying to create stimuli that represent only one form of relatedness, we included a range of each of these variables to explore their potential interaction.

Specifically, this research was conceptualized within the framework of a three-tiered view of the interconnections between these systems as it relates to processing concept information. The three-tiered view is inspired by models of reading and naming, particularly

163   the triangle models presented by Seidenberg and McClelland (1989) and Plaut (1995). These

164   models explored the nature of reading as bidirectional relations between semantics,

165   orthography, and phonology. In this research, we examine if the associative, semantic, and

166   thematic systems are interactive for judgment and recall processes, much like the proposed

167   interactive nature of phonology, orthographics, and semantics for reading and naming

168   processes. Potentially, association, semantic, and thematic facets of word relation each

169   provide a unique contribution that can be judged and used for memory, thus, suggesting

170   three separate networks of independent information. This view seems unlikely, in that

171   research indicates that there is often overlap in the information provided by each measure of

172   word-pair relatedness. Instead, dynamic attractor networks, as proposed by Hopfield (1982)

173   and McLeod, Shallice, and Plaut (2000) may better represent the interplay between these

174   representations of concepts, as these models posit a similar feedback relationship between

175   concepts in a network. Using these models as a theoretical framework for our study, we

176   sought to understand how these three types of word-pair information may interact when

177   judgment and recall processes were applied to concept networks, and use it as a framework

178   for exploring how associative, semantic, and thematic memory networks share

179   interconnections. Therefore, this study provides evidence of the structure and interplay

180   between different forms of network relations for two cognitive tasks of judgment and retrieval

181   and will shed light on the underlying processing for each task.

182   ##Application to Judgment Studies

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

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

185   they would be able to correctly respond with the target with the cue on a recall task.

186   Judgments are typically made out of 100, with a participant response of 100 indicating full

187   confidence in recall ability. In their 2005 study, Koriat and Bjork examined overconfidence in

188   JOLs by manipulating associative relations (FSG) between word-pairs and found that

subjects were more likely to overestimate recall for pairs with little or no associative relatedness. Additionally, this study found that when accounting for associative direction, subjects were more likely to overestimate recall for pairs that were high in backwards strength but low in forward strength. To account for this finding, the authors suggested that JOLs may rely more heavily on overlap between cue and target with the direction of the associative relationship being secondary. Take for example the pair *feather - bird*, which has a FSG of .051 and a BSG of .359. This item pair also has a cosine value of .272 (suggesting low to moderate feature overlap) and an LSA score of .517 (suggesting moderate thematic overlap). As such, some of the overconfidence in JOLs may be attributed more than just item associations. Paired items may also be connected by similar themes or share certain features, resulting in inflated JOLs.

Expanding upon this research, the traditional judgment of learning task (JOL) can be manipulated to investigate perceptions of word pair relationships by having participants judge how related they believe the cue and target items to be (Maki, 2007a, 2007b). The judged values generated from this task can then be compared to the normed databases to create a similar accuracy function or correlation as is created in JOL studies. When presented with the item pair, participants are asked to estimate the number of people out of 100 who would provide the target word when shown only the cue (Maki, 2007b), which mimics how the association word norms are created through free association tasks. Maki (2007a) investigated such judgments within the context of associative memory by having participants rate how much associative overlap was shared between items and found that responses greatly overestimated the actual overlap strength for pairs that were weak associates, while underestimating strong associates; thus replicating the Koriat and Bjork (2005) findings for relatedness judgments based upon associative memory, rather than judgments based on learning.

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

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

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

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

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

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

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

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

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

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

memory with the same results.

The present study combined the paradigms of PAL, JOLs, and JAM to examine item

recall and judgments for three types of judgments of relatedness (JORs) to explore the

underlying memory network that is used for each of these cognitive processes as described

above. We tested the following hypotheses based on previous research and semantic memory

models:

1) First, we sought to expand previous Maki (2007b), Maki (2007a), Buchanan (2010),

and Valentine and Buchanan (2013) research to include three types of JORs 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 JOR

and calculated average slope and intercept values for each participant. First, we

expected to find slope and intercept values that were significantly different from zero.

Though the three types of word relations are distinct from one another, we should

expect to find slopes and intercepts for semantic and thematic JORs to be within the

range of previous JAM findings if these memory systems are interconnected. Finally,

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

241   associative JORs, etc.). This hypothesis explores if judgment findings replicate across a

242   range of variables and covariates (rather than each individually, as previous JOL and

243   JAM publications) and expands our knowledge on how the judgment process taps into

244   the underlying memory network.

245   2) Next, we explored the predictions from semantic network models that the relation

246   between association, semantics, and thematics would be bidirectional in nature (i.e.,

247   the three-tiered hypothesis of each type of knowledge stacked in memory). Therefore,

248   we expected to find an interaction between database norms in predicting JORs. We

249   used multilevel modeling to examine the interaction of database norms for association,

250   semantics, and thematics in relation to participant judgments.

251   3) These analyses were then extended to recall as the dependent variable of interest. We

252   tested for the interaction of database norms in predicting recall by using a multilevel

253   logistic regression, while controlling for judgment condition and rating. We expected to

254   find that database norms would show differences in recall based on the levels of other

255   variables (the interaction would be significant), and that ratings would also positively

256   predict recall (i.e., words that participants thought were more related would be

257   remembered better). Because judgment and recall are different cognitive processes, we

258   used this hypothesis to examine how memory networks may be differently interactive

259   for memory in comparison to judgment.

260   4) Finally, we examined if the judgment slopes from Hypothesis 1 would be predictive of

261   recall. Hypothesis 3 examined the direct relationship of word relatedness on recall,

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

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

264   for multiple judgment slope conditions. This hypothesis combines both cognitive

265   processes into one analysis, to explore how judgment ability (i.e., slopes) would impact

266   the memory process.

267   **Method**

268   **Participants**

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

270   2016). This package uses simulations to generate power estimates for mixed linear models

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

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

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

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

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

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

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

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

279   Responses were rejected for participants who entered related words when numerical

280   judgment responses were required, and for participants who responded to the cue words

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

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

283   ##Materials

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

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

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

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

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

289   indicates that participants would always give a target word in response to the cue word.

290   Semantic relatedness was measured with cosine (COS), which is a measure of semantic

291   feature overlap (Buchanan et al., 2013; McRae et al., 2005; Vinson & Vigliocco, 2008). This

292 variable ranges from zero to one where zero indicates no shared semantic features between

293 concepts and higher numbers indicate more shared features between concepts. Thematic

294 relatedness was calculated with Latent Semantic Analysis (LSA), which generates a score

295 based upon the co-occurrences of words within a document (Landauer & Dumais, 1997;

296 Landauer et al., 1998). LSA values also range from zero to one, indicates no co-occurrence at

297 the low end and higher co-occurrence with higher values. These values were chosen to

298 represent these categories based on face validity and previous research on how word pair

299 psycholinguistic variables overlap (Maki & Buchanan, 2008).

300   The selected stimuli included a range of values for each variable. Table 1 displays

301 stimuli averages, SD, and ranges. A complete list of stimuli can be found at

302 http://osf.io/y8h7v. The stimuli were arranged into three blocks for each judgment

303 condition described below wherein each block contained 21 word pairs. Due to limitations of

304 the available stimuli, blocks were structured so that each one contained seven word pairs of

305 low (0-.33), medium (.34-.66), and high (.67-1.00) COS relatedness. Because of this selection

306 process, FSG and LSA strengths are contingent upon the selected stimuli's COS strengths.

307 We selected stimuli within the cosine groupings to cover a range of FSG and LSA values, but

308 certain combinations are often difficult to achieve. For example, there are only four

309 word-pairs that are both high COS and high FSG, thus limiting the ability to manipulate

310 LSA. The study was built online using Qualtrics, and three surveys were created to

311 counter-balance the order in which judgment conditions appeared. Each word pair appeared

312 counter-balanced across each judgment condition, and stimuli were randomized within each

313 block.

314 **Procedure**

315   The present study was divided into three phases. In the first phase, JORs were elicited

316 by presenting participants with word pairs and asking them to make judgments of how

₃₁₇ related they believed the words in each pair to be. This judgment phase consisted of three

₃₁₈ blocks of 21 word pairs which corresponded to one of three types of word pair relationships:

₃₁₉ associative, semantic, or thematic. Each block was preceded by a set of instructions

₃₂₀ explaining one of the three types of relationships, and participants were provided with

₃₂₁ examples which illustrated the type of relationship to be judged. Participants were then

₃₂₂ presented with the word pairs to be judged. The associative block began by explaining

₃₂₃ associative memory and the role of free association tasks. Participants were provided with

₃₂₄ examples of both strong and weak associates. For example, *lost* and *found* and were

₃₂₅ presented as an example of a strongly associated pair, while *article* was paired with

₃₂₆ *newspaper*, *the*, and *clothing* to illustrate that words can have many weak associates. The

₃₂₇ semantic judgment block provided participants with a brief overview of how words are

₃₂₈ related by meaning and showed examples of concepts with both high and low feature overlap.

₃₂₉ *Tortoise* and *turtle* were provided as an example of two concepts with significant overlap.

₃₃₀ Other examples were then provided to illustrate concepts with little or no overlap. For the

₃₃₁ thematic judgments, participants were provided with an explanation of thematic relatedness.

₃₃₂ *Tree* is explained to be related to *leaf*, *fruit*, and *branch*, but not *computer*. Participants were

₃₃₃ then given three concepts (*lost*, *old*, *article*) and were asked to come up with words that they

₃₃₄ feel are thematically related.

₃₃₅ After viewing the examples at the start of the block, participants completed the JOR

₃₃₆ task. Each block contained a set of instructions which were contingent upon the type of JOR

₃₃₇ being elicited. For example, instructions in the associative block asked participants to

₃₃₈ estimate how many individuals out of 100 they expect would respond to the cue word with a

₃₃₉ given target, instructions for semantic JORs asked participants to indicate the percent of

₃₄₀ features shared between two concepts, and instructions for the thematic JOR task asked

₃₄₁ participants to base ratings on how likely to words would be used together in the same story.

₃₄₂ The complete experiment can be found at http://osf.io/y8h7v, which contains the exact

₃₄₃ instructions given to participants for each block and displays the structure of the study. All

344 instructions were modeled after Buchanan (2010) and Valentine and Buchanan (2013).

345 In accordance with previous work on JOLs and JAM, participants made JOR ratings
346 using a scale of zero to one hundred, with zero indicating no relationship, and one hundred
347 indicating a perfect relationship. Participants typed their responses into the survey. Once
348 completed, participants then completed the remaining judgment blocks in the same manner.
349 Each subsequent judgment block changed the type of JOR being made. Three versions of
350 the study were created, which counter-balanced the order in which the judgment blocks
351 appeared, and participants were randomly assigned to a survey version. This resulted in each
352 word pair receiving a relatedness judgments on each of the three types relationships.

353 After completing the judgment phase, participants were then presented with a short
354 distractor task to account for recency effects. In this section, participants were presented
355 with a randomized list of the fifty U.S. states and were asked to arrange them in alphabetical
356 order. This task was timed to last two minutes. Once time had elapsed, participants
357 automatically progressed to the final phase, which consisted of a cued-recall task.
358 Participants were presented with each of the 63 cue words from the judgment phase and
359 were asked to complete each word pair by responding with the correct target word.
360 Participants were informed that they would not be penalized for guessing. The cued-recall
361 task included all stimuli in a random order.

362 **Results**

363 **Data Processing and Descriptive Statistics**

364 First, the results from the recall phase of the study was coded as zero for incorrect
365 responses, one for correct responses, and NA for participants who did not complete the recall
366 section (all or nearly all responses were blank). All word responses to judgment items were
367 deleted and set to missing data. The final dataset was created by splitting the initial data

368 file into six sections (one for each of the three experimental blocks and their corresponding

369 recall scores). Each section was individually melted using the *reshape* package in *R*

370 (Wickham, 2007) and was written as a csv file. The six output files were then combined to

371 form the final dataset. Code is available on our OSF page embedded inline with the

372 manuscript in an *R* markdown document written with the *papaja* package (Aust & Barth,

373 2017). With 112 participants, the dataset in long format included 7,056 rows of potential

374 data (i.e., 112 participants * 63 JORs). One out of range JOR data point (> 100) was

375 corrected to NA. Missing data for JORs or recall were then excluded from the analysis,

376 which included word responses to judgment items (i.e., responding with *cat* instead of a

377 number). These items usually excluded a participant from receiving Amazon Mechanical

378 Turk payment, but were included in the datasets found online. In total, 787 data points were

379 excluded (188 JOR only, 279 recall only, 320 both), leading to a final *N* of 105 participants

380 and 6,269 observations. Recall and JOR values were then screened for outliers using

381 Mahalanobis distance at $p < .001$, and no outliers were found (Tabachnick & Fidell, 2012).

382 To screen for multicollinearity, we examined correlations between judgment items, COS,

383 LSA, and FSG. All correlations were rs $< .50$.

384   The mean JOR for the associative condition ($M = 58.74$, $SD = 30.28$) was lower than

385 the semantic ($M = 66.98$, $SD = 28.31$) and thematic ($M = 71.96$, $SD = 27.80$) conditions.

386 Recall averaged over 60% for all three conditions: associative $M = 63.40$, $SD = 48.18$;

387 semantic $M = 68.02$, $SD = 46.65$; thematic $M = 64.89$, $SD = 47.74$.

Hypothesis 1

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

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

391 memory. FSG, COS, and LSA were used to predict each type of relatedness judgment. JOR

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

₃₉₃ Slopes and intercepts were then calculated for each participant's ratings for each of the three

₃₉₄ JOR conditions, as long as they contained at least nine data points out of the twenty-one

₃₉₅ that were possible. Single sample t-tests were then conducted to test if slope and intercept

₃₉₆ values significantly differed from zero. See Table 2 for means and standard deviations. Slopes

₃₉₇ were then compared to the JAM function, which is characterized by high intercepts (between

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

₃₉₉ scaling of our data, to replicate this function, we should expect to find intercepts ranging

₄₀₀ from .40 to .60 and slopes in the range of 0.20. to 0.40. Intercepts for associative, semantic,

₄₀₁ and thematic JORs were each significant, and all fell within or near the expected range.

₄₀₂ Overall, thematic JORs had the highest intercept at .656, while JORs elicited in the

₄₀₃ associative condition had the lowest intercept at .511.

₄₀₄   The JAM slope was successfully replicated for FSG in the associative JOR condition,

₄₀₅ with FSG significantly predicting association, although the slope was slightly higher than

₄₀₆ expected at .491. COS and LSA did not significantly predict association. For semantic

₄₀₇ judgments, each of the three database norms were significant predictors. However, JAM

₄₀₈ slopes were not replicated for this judgment type, as FSG had the highest slope at .118,

₄₀₉ followed by LSA .085, and then COS .059. These findings were mirrored for thematic JORs,

₄₁₀ as each database norm was a significant predictor, yet slopes for each predictor fell below

₄₁₁ range of the expected JAM slopes. Again, FSG had the highest slope, this time just out of

₄₁₂ range at .192, followed closely by LSA at .188. Interestingly, COS slopes were found to be

₄₁₃ negative for this judgment condition, -.081. Overall, although JAM slopes were not

₄₁₄ successfully replicated in each JOR condition, the high intercepts and shallow slopes present

₄₁₅ across conditions are still indicative of overconfidence and insensitivity in participant JORs.

₄₁₆   Additionally, we examined the frequency that each predictor variable was the strongest

₄₁₇ predictor for each of the three JOR conditions. For the associative condition, FSG was the

₄₁₈ strongest predictor for 64.0% of the participants, with COS and LSA being the strongest for

only 16.0% and 20.0% of participants respectively. These differences were less distinct when examining the semantic and thematic JOR 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 condition. Therefore, these results provide evidence of the nature of judgments on the memory network as each judgment type appeared to tap each tier differently, suggesting a three-part system, rather than one large, encompassing memory network.

Hypothesis 2

The goal of Hypothesis 2 was to test for an interaction between the three database norms when predicting participant JORs to examine the bidirectional network model. 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, & Team, 2017). A maximum likelihood multilevel model was used to test the interaction between FSG, COS, and LSA when predicting JOR values, with participant number used as the random intercept factor. The type of JOR being elicited was controlled for, so as to better assess the impact of each word overlap measure regardless of JOR condition. Multilevel models were used to retain all data points (rather than averaging over items and conditions) while controlling for correlated error due to participants, which makes these models 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 (β = -1.492, $p < .001$), average COS (β =-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 JOR values ($\beta = 0.663$, p < .001). At low COS and average LSA, FSG decreased but still significantly predicted JORs ($\beta = 0.375$, 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 JORs. 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 JORs ($\beta = 0.167$, p < .001), and finally when both COS and LSA were high, FSG increased and was a significant predictor of JOR values ($\beta = 0.236$, p < .001). Thus, at high levels of semantic overlap, associative and thematic overlap are complementary when predicting JOR ratings, increasing together as semantic strength increases. Figure 2 displays the three-way interaction wherein the top row of figures indicates the seesaw effect, as thematic strength increases, the predictive ability of associative overlap decreases in strength. The bottom row indicates the complementary effect where increases in LSA occur with increases in FSG predictor strength. Therefore, the cognitive process of judgment appears to be interactive in nature across these three types of

471 memory information.

472 **Hypothesis 3**

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

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

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

476 predicting participant recall. As with the previous hypothesis, we controlled for JOR

477 condition and, additionally, covaried JOR ratings. Participants were used as a random

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

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

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

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

482   The same moderation process used in Hypothesis 2 was then repeated, with simple

483 slopes first calculated at low, average, and high levels of COS. This set of analyses resulted

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

485 and high COS ($\beta = 5.811$, $p = .009$). No significant two-way interaction was found at

486 average COS ($\beta = -1.017$, $p = .493$). Following the design of hypothesis two, simple slopes

487 were then calculated for low, average, and high levels of LSA at the low and high levels of

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

489 When both COS and LSA were low, FSG was a significant predictor of recall ($\beta = 4.116$, p

490 $< .001$). At low COS and average LSA, FSG decreased from both low levels, but was still a

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

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

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

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

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

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

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

$= 4.692$, $p < .001$). Therefore, at high levels of at high levels of semantics, thematics and

association are complementary predictors of recall, increasing together and extending the

findings of Hypothesis 2 to participant recall. Figure 3 displays the three-way interaction.

The top left figure indicates the counterbalancing effect of recall of LSA and FSG, while the

top right figure shows no differences in simple slopes for average levels of cosine. The bottom

left figure indicates the complementary effects where LSA and FSG increase together as

predictors of recall at high COS levels.

Hypothesis 4

In our fourth and final hypothesis, we investigated whether the JOR 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 condition, so that each set of JOR 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 JOR 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

$_{521}$ ($b = 0.501$, $p = .279$) were non-significant. In the semantic condition, COS slope ($b = 2.039$,

$_{522}$ $p < .001$) and LSA slope ($b = 1.061$, $p = .020$) were both found to be significant predictors

$_{523}$ of recall. FSG slope was non-significant in this condition ($b = 0.381$, $p = .187$). Finally, no

$_{524}$ predictors were significant in the thematic condition, though LSA slope was found to be the

$_{525}$ strongest ($b = 0.896$, $p = .090$). This analysis indicated the extent to which the cognitive

$_{526}$ processes are related to each other as part of the memory network (i.e., judgment sensitivity

$_{527}$ predicting recall), furthering Hypothesis 2 and 3 which illustrated the nature of those

$_{528}$ cognitive processes' relationship with the underlying memory network.

$_{529}$ <div align="center">**Discussion**</div>

$_{530}$ This study investigated the relationship between associative, semantic, and thematic

$_{531}$ word relations and their effect on participant JORs and recall performance through the

$_{532}$ testing of four hypotheses. In our first hypothesis, bias and sensitivity findings first proposed

$_{533}$ by Maki (2007a) were successfully replicated in the associative condition, with slope and

$_{534}$ intercept values falling within the expected range. While these findings were not fully

$_{535}$ replicated when extending the analysis to include semantic and thematic JORs (as slopes in

$_{536}$ these conditions did not fall within the appropriate range), participants still displayed high

$_{537}$ intercepts and shallow slopes, suggesting overconfidence in judgment making and an

$_{538}$ insensitivity to changes in strength between pairs. Additionally, when looking at the

$_{539}$ frequency that each predictor was the strongest in making JORs, FSG was the best predictor

$_{540}$ for both the associative and semantic conditions, while LSA was the best predictor in the

$_{541}$ thematic condition. In each of the three conditions, COS was the weakest predictor, even

$_{542}$ when participants were asked to make semantic judgments. This finding suggests that

$_{543}$ associative relationships seem to take precedence over semantic relationships when judging

$_{544}$ pair relatedness, regardless of what type of JOR is being elicited. Additionally, this finding

$_{545}$ may be taken as further evidence of a separation between associative information and

546 semantic information, in which associative information is always processed, while semantic

547 information may be suppressed due to task demands (Buchanan, 2010; Hutchison & Bosco,

548 2007).

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

550 LSA when predicting participant JORs. At low semantic overlap, a seesaw effect was found

551 in which increases in thematic strength led to decreases in associative predictiveness. This

552 finding was then replicated in Hypothesis 3 when extending the analysis to predict recall. By

553 limiting the semantic relationships between pairs, an increased importance is placed on the

554 role of associations and thematics when making relatedness judgments or retrieving pairs. In

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

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

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

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

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

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

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

562 boosts the predictive ability of associative word norms for both recall and JORs.

563   Finally, our fourth hypothesis used the JOR slopes and intercepts calculated in

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

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

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

567 LSA slopes, with COS being the strongest. However, for the thematic condition, although

568 the LSA slope was the strongest, no predictors were significant. One explanation for this

569 finding is that thematic relationships between item pairs act as a blend between associations

570 and semantics. As such, LSA faces increased competition from the associative and semantic

571 database norms when predicting recall in this manner. Additionally, the dominance of FSG

572 when predicting recall in the associative condition may be attributed to word associations

573 being more accessible (and, thus, easier to process) than semantic or thematic relations

574 between pairs.

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

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

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

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

579 handle storage and processing for meaning and association (see Ferrand & New, 2004 for a

580 review), the presence of these interactions suggests that connections exist between these

581 individual memory networks, linking them to one another. As such, we suggest that these

582 memory systems may be connected in such a way to form a three-tiered, interconnected

583 system. First, information enters the semantic memory network, which processes features of

584 concepts and provides a means of categorizing items based on the similarity of their features.

585 Next, the associative network adds information for items based on contexts generated by

586 reading or speech. Finally, the thematic network pulls in information from both the semantic

587 and associative networks to create a mental representation of both the item and its place in

588 the world relative to other concepts. This study did not explore the timing of information

589 input from each of these systems, but it may be similar to a dual-route model of reading and

590 naming, in that each runs in parallel contributing the judgment and recall process

591 (Coltheart, Curtis, Atkins, & Haller, 1993).

592   Viewing this model purely through the lens of semantic memory, it draws comparison

593 to dynamic attractor models (Hopfield, 1982; Jones et al., 2015; McLeod et al., 2000). One

594 of the defining features of dynamic attractor models is that they allow for some type of

595 bidirectionally or feedback between connections in the network. In the study of semantic

596 memory, these models are useful for taking into account multiple restraints such as links

597 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 class of models are those based upon Seidenberg and McClelland (1989) "triangle model" (see Harley, 2008 for a review). They key feature of these models is that they recognize speech and reading based upon the orthography, phonology, and meaning of words in a bidirectional manner, similar to the models described above. Harm and Seidenberg (2004) developed a version which included a focus on semantics, with word meaning being related to input from the orthography and phonology components of the model. Our findings from the present study further suggest that thematic and associative knowledge is incorporated with meaning. One way of framing our results within this literature is to consider the semantic section of the triangle model as being comprised of these three tiers, and that concept information is processed to some degree on each of these domains. One area for future studies of this nature may be investigating how aspects of orthography and phonology impact these memory networks. Additionally, future studies may wish to consider elements of thematic and associative knowledge when examining semantic based tasks, such as word recognition and reading, as thematic and associative information is interconnected with the semantic network. Ultimately, further studies will be needed to fully understand the interconnections between the semantic, thematic, and associative networks.

Compliance with Ethical Standards

The authors declare that they have no conflict of interest. The study was approved by the Institutional Review Board at Missouri State University. Particiants filled out an

623 informed consent at the beginning of the study, after accepting the HIT on Mechanical Turk.

624 The complete study with consent form can be found on our OSF page: http://osf.io/y8h7v.

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

De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2018). *Measuring the associative structure of English: The "Small World of Words" norms for word association.*

De Deyne, S., Navarro, D. J., & Storms, G. (2013). Better explanations of lexical and

647   semantic cognition using networks derived from continued rather than single-word

648   associations. *Behavior Research Methods*, *45*(2), 480–498.

649   doi:10.3758/s13428-012-0260-7

650   Ferrand, L., & New, B. (2004). Semantic and associative priming in the mental lexicon. In P.

651   Bonin (Ed.), *The mental lexicon* (pp. 25–43). Hauppauge, NY: Nova Science.

652   Gelman, A. (2006). Multilevel (hierarchical) modeling: What it can and cannot do.

653   *Technometrics*, *48*(3), 432–435. doi:10.1198/004017005000000661

654   Green, P., & MacLeod, C. J. (2016). SIMR: an R package for power analysis of generalized

655   linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498.

656   doi:10.1111/2041-210X.12504

657   Harley, T. (2008). *The psychology of language: From data to theory* (Third.). New York:

658   Psychology Press.

659   Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading:

660   Cooperative division of labor between visual and phonological processes.

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

662   Hertzog, C., Kidder, D. P., Powell-Moman, A., & Dunlosky, J. (2002). Aging and monitoring

663   associative learning: Is monitoring accuracy spared or impaired? *Psychology and*

664   *Aging*, *17*(2), 209–225. doi:10.1037/0882-7974.17.2.209

665   Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective

666   computational abilities. *Proceedings of the National Academy of Sciences*, *79*(8),

667   2554–2558. doi:10.1073/pnas.79.8.2554

668   Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap?

669   A microanalytic review. *Psychonomic Bulletin & Review*, *10*(4), 785–813.

670   doi:10.3758/BF03196544

671   Hutchison, K. A., & Bosco, F. A. (2007). Congruency effects in the letter search task:

672   Semantic activation in the absence of priming. *Memory & Cognition, 35*(3), 514–525.

673   doi:10.3758/BF03193291

674   Jiang, J. J., & Conrath, D. W. (1997). Semantic similarity based on corpus statistics and

675   lexical taxonomy. *Proceedings of International Conference Research on Computational*

676   *Linguistics (ROCLING X)*. Retrieved from http://arxiv.org/abs/cmp-lg/9709008

677   Jones, L. L., & Golonka, S. (2012). Different influences on lexical priming for integrative,

678   thematic, and taxonomic relations. *Frontiers in Human Neuroscience, 6*(July), 1–17.

679   doi:10.3389/fnhum.2012.00205

680   Jones, M. N., Willits, J., & Dennis, S. (2015). Models of semantic memory. In A. T.

681   Townsend & J. R. Busemeyer (Eds.), *Oxford handbook of mathematical and*

682   *computational psychology* (pp. 232–254). Oxford University Press.

683   doi:10.1093/oxfordhb/9780199957996.013.11

684   Jouravlev, O., & McRae, K. (2016). Thematic relatedness production norms for 100 object

685   concepts. *Behavior Research Methods, 48*(4), 1349–1357.

686   doi:10.3758/s13428-015-0679-8

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

688   during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition,*

689   *31*(2), 187–194. doi:10.1037/0278-7393.31.2.187

690   Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent

691   semantic analysis theory of acquisition, induction, and representation of knowledge.

692   *Psychological Review, 104*(2), 211–240. doi:10.1037//0033-295X.104.2.211

Landauer, T. K., Foltz, P. W., Laham, D., Folt, 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

716   production norms for a large set of living and nonliving things. *Behavior Research*

717   *Methods*, *37*(4), 547–559. doi:10.3758/BF03192726

718   Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words:

719   Evidence of a dependence between retrieval operations. *Journal of Experimental*

720   *Psychology*, *90*(2), 227–234. doi:10.1037/h0031564

721   Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1975). Loci of contextual effects on

722   visual word-recognition. In P. M. A. Rabbitt (Ed.), *Attention and performance v.*

723   London, UK: Academic Press.

724   Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*,

725   *38*(11), 39–41. doi:10.1145/219717.219748

726   Nelson, D. L., McEvoy, C. L., & Dennis, S. (2000). What is free association and what does it

727   measure? *Memory & Cognition*, *28*(6), 887–899. doi:10.3758/BF03209337

728   Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida

729   free association, rhyme, and word fragment norms. *Behavior Research Methods,*

730   *Instruments, & Computers*, *36*(3), 402–407. doi:10.3758/BF03195588

731   Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*,

732   *76*(3), 241–263. doi:10.1037/h0027272

733   Pinheiro, J., Bates, D., Debroy, S., Sarkar, D., & Team, R. C. (2017). nlme: Linear and

734   nonlinear mixed effects models. Retrieved from

735   https://cran.r-project.org/package=nlme

736   Plaut, D. C. (1995). Semantic and associative priming in a distributed attractor network.

737   *Proceedings of the 17th Annual Conference of the Cognitive Science Society*, 37–42.

738   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., & PDP Research Group. (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*(4), 523–568. doi:10.1037//0033-295X.96.4.523

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. (2012). *Using multivariate statistics* (Sixth.). Boston, MA: Pearson.

Valentine, K. D., & Buchanan, E. M. (2013). JAM-boree: An application of observation oriented modelling to judgements of associative memory. *Journal of Cognitive*

762   *Psychology, 25*(4), 400–422. doi:10.1080/20445911.2013.775120

763   Vinson, D. P., & Vigliocco, G. (2008). Semantic feature production norms for a large set of

764   objects and events. *Behavior Research Methods, 40*(1), 183–190.

765   doi:10.3758/BRM.40.1.183

766   Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical*

767   *Software, 21*(12), 1–20. 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.
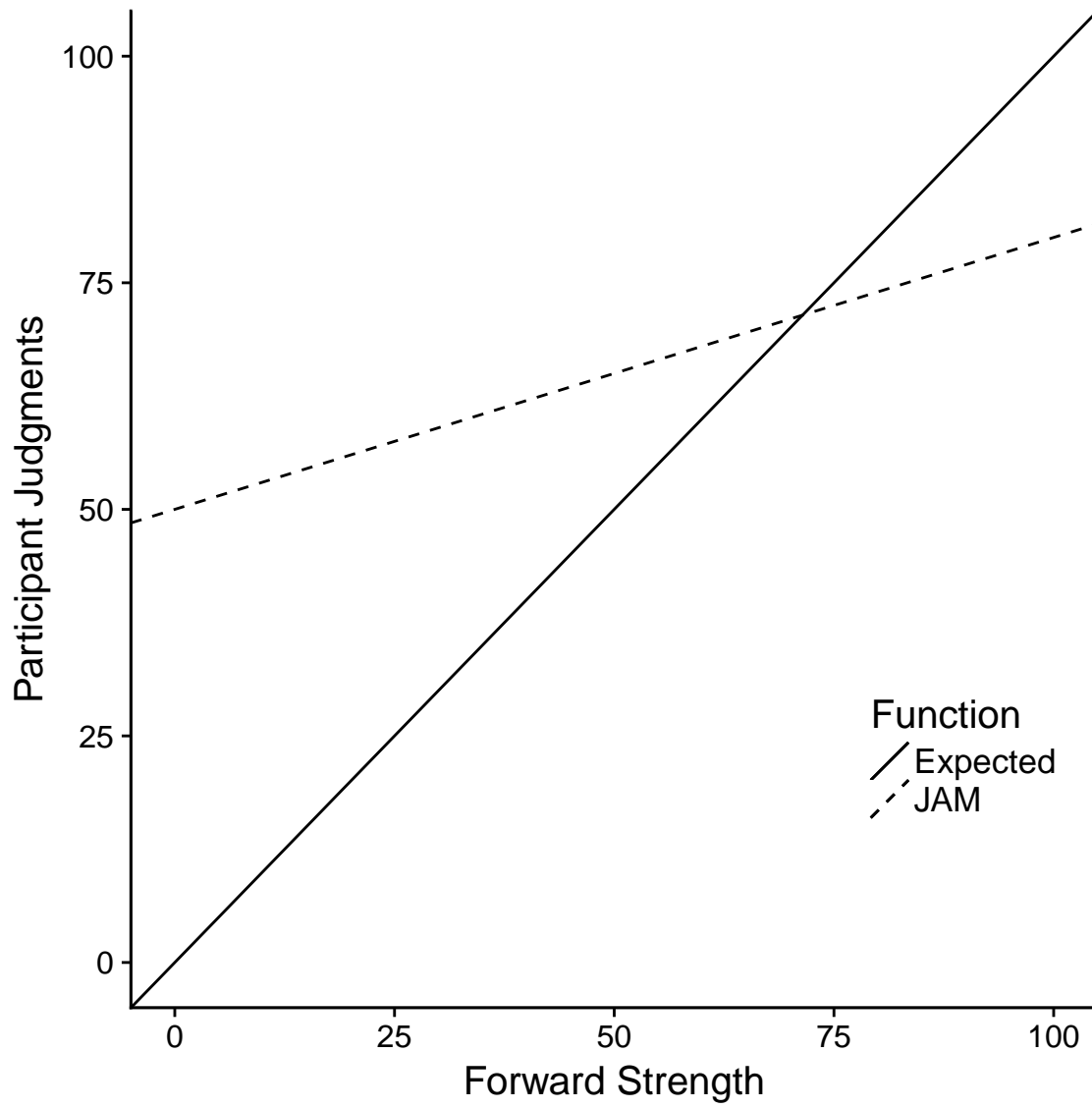
Figure 1. JAM slope findings from Maki (2007a). JAM is characterized by a high intercept (between 40 and 60) and a shallow slope (between 0.20 and 0.40). The solid line shows expected results if judgment ratings are perfectly calibrated with association norms.
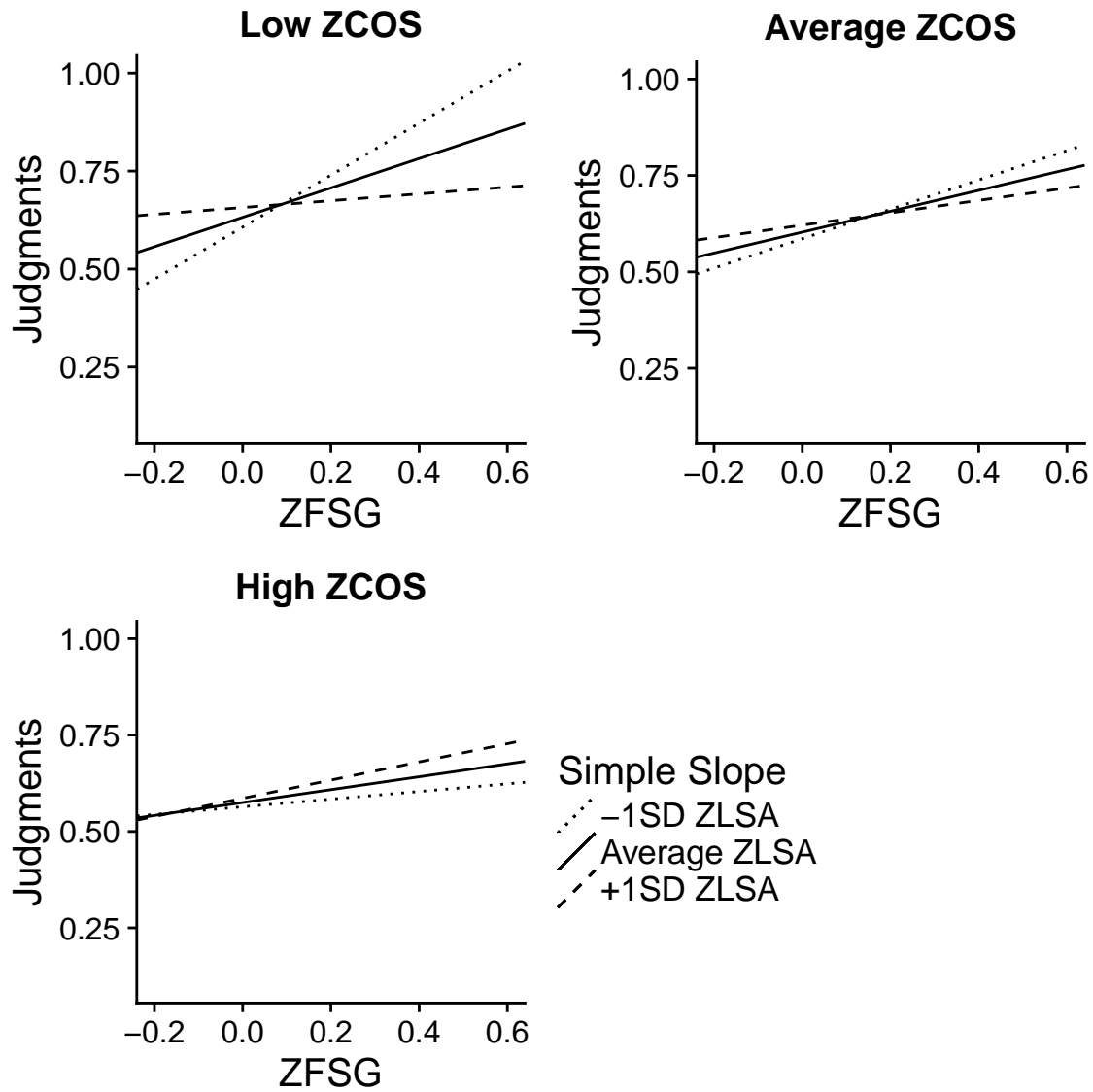
Figure 2. Simple slopes graph displaying the slope of FSG when predicting JORs at low, average, and high LSA split by low, average, and high COS. All variables were mean centered.
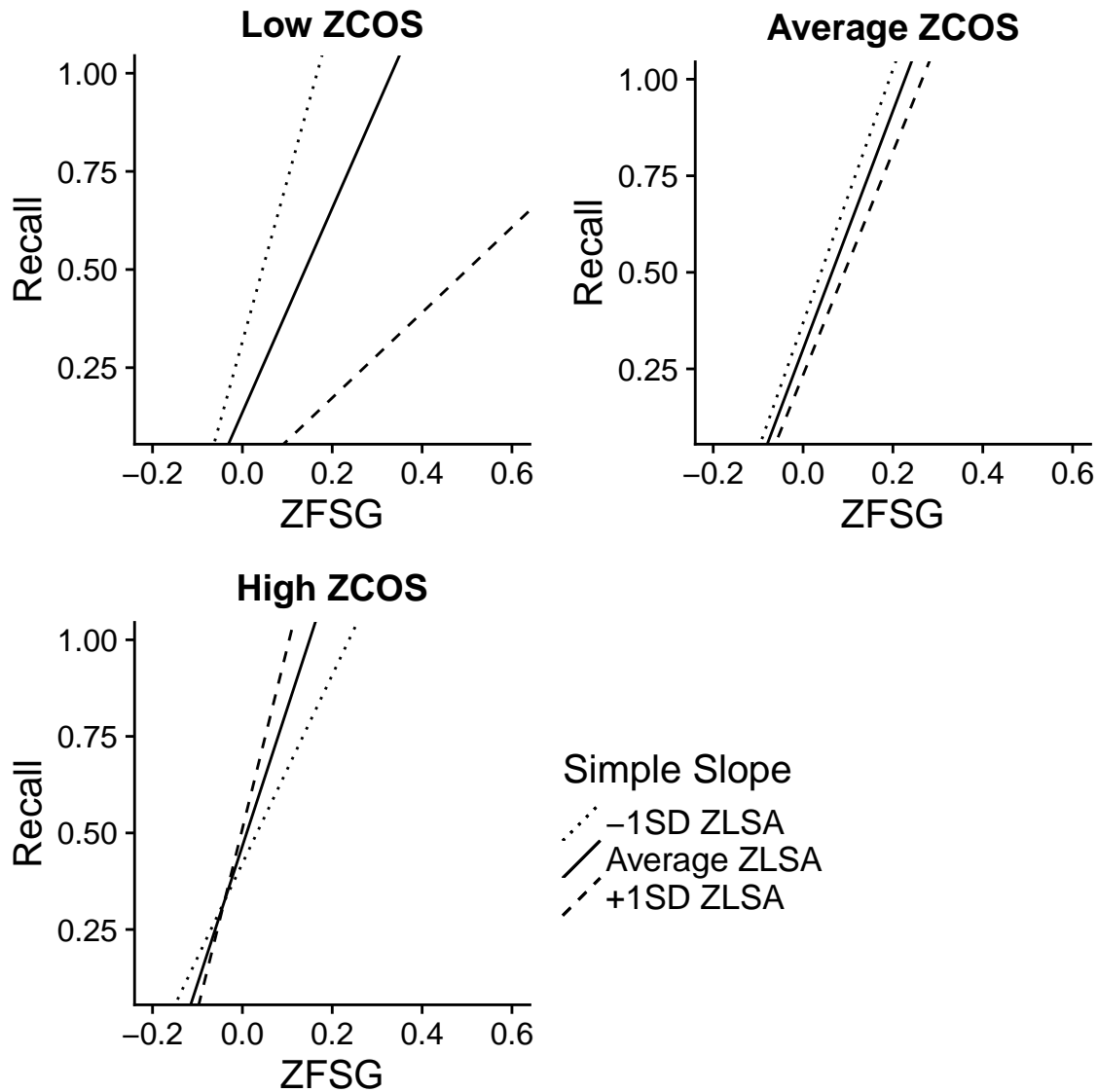
Figure 3. Simple slopes graph displaying the slope of FSG when predicting recall at low, average, and high LSA split by low, average, and high COS. All variables were mean centered.