Investigating the Interaction of Direct and Indirect Relation on Memory Judgments and Retrieval

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

[1] University of Southern Mississippi

[2] Harrisburg University of Science and Technology

Author Note

Nicholas P. Maxwell is a Ph.D. candidate at the University of Southern Mississippi. Erin M. Buchanan is an Professor of Cognitive Analytics at Harrisburg University of Science and Technology.

Correspondence concerning this article should be addressed to Erin M. Buchanan, 326 Market St., Harrisburg, PA, 17101. E-mail: ebuchanan@harrisburgu.edu

Abstract

This study examined the interactive relationship between direct and indirect relation strength in the prediction of item judgments and cued-recall performance. Participants were recruited from Amazon's Mechanical Turk and were given word pairs of varying relatedness to judge for their semantic, thematic, and associative strength. After completing a distractor task, participants then completed a cued recall task. First, we sought to expand previous work on judgments of associative memory (JAM) to include semantic and thematic based judgments, while also replicating bias and sensitivity findings. Next, we tested for an interaction between direct and indirect association when predicting participant judgments while also expanding upon previous work by examining that interaction when predicting recall. The interaction between direct and indirect association was significant for both judgments and recall. For low indirect association, direct association was the primary predictor of both judgment strength and recall proportions. However, this trend reversed for high indirect association, as higher levels of indirect relation decreased the effectiveness of direct relation as a predictor. Overall, our findings indicate the degree to which the processing of associative, semantic, and thematic information impacts cognitive processes such as retrieval and item judgments, while also examining the underlying, interactive relationship that exists in language used to represent concept information.

Keywords: judgments, memory, association, semantics, thematics

Investigating the Interaction of Direct and Indirect Relation on Memory Judgments and Retrieval

The study of cognition has a rich history of exploring the way in which associations affect human memory. One key finding is that associations between items influence cognitive processing and 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, & Dennis, 2000).

Measuring Association

Within cognitive psychology, word associations have been conceptualized differently across various lines of research (i.e., direct word associations, mediated associates, etc.; see De Deyne et al. (2013b) for a review). For the present study, we focus on two types of associations: direct associations and indirect associations. Direct word associations are traditionally viewed as the context-based relation between concepts, usually found in text or popular culture (Nelson et al., 2000). Within this framework, word associations are thought to arise in several different ways. Such associations may develop through their co-occurrence together in either written or spoken language. 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 overlap in terms of meaning. However, this separation is not the case for all associative pairs. For example, word associations capture the knowledge that fish live in water (e.g., *fish* – *swim*) and that dogs and cats share many similar features. To generate norms measuring direct associations, 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 (i.e., the pair's normed forward strength, FSG) 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 (Nelson et al., 2000). Thus, the free association process can be thought of as generating an index that contains the relative accessibility of related words in memory (Nelson, McEvoy, & Schreiber, 2004).

Using this technique, researchers have developed databases of associative word norms that can be used to generate stimuli, generally with a high degree of reliability (e.g., The University of South Florida Free Association Norms; Nelson et al., 2004). However, this reliability becomes questionable for weak associates. Because the traditional free association task focuses only the first word that is provided the cue, target items that are more weakly associated may become underrepresented in the dataset, as the inclination to respond with stronger associates may disrupt access to weaker associates (i.e., the availability heuristic]. Recently, The Small World of Words project (SWOW, De Deyne et al., 2013b; De Deyne, Navarro, Perfors, Brysbaert, & Storms, 2018) has sought to correct for this sampling issue by employing a multiple response free association task. In this modified free association task, subjects are asked to generate three target items in response to the cue. The updated SWOW association norms provide several advantages when compared to other collections of free association norms. First, this norm set is the largest to date, consisting of approximately 12,000 cue items (for comparison, the USF norms consist of 5,400 cue items). Because of its large size, the SWOW norms provide a better approximation of natural language. Second, the use of a multiple response technique allows for greater reliability of weak associates, resulting in more weak associations being captured by the network, as weak associates are rarely given as the first response and thus may be under represented when only one response is elicited (De Deyne et al., 2013b).

Measuring Relatedness

Whereas direct associations focus on the relationships between individual words, indirect associations focus on how a concept fits into the overall structure of the semantic network (De Deyne et al., 2013b; Deese, 1965). Because indirect associations capture information derived from the overall structure of the semantic network, these norms can also be used to represent semantic properties of item pairs and can be used to approximate links between concepts within semantic memory networks. This includes mediated associates (i.e., *lion – stripes* is mediated through *tiger*; see Huff & Hutchison, 2011 for a review of mediated associates) and is one of the underlying factors behind distributional models of semantic memory (e.g., Latent Semantic Analysis, Landauer & Dumais, 1997; Hyperspace Analogue to Language Model, Lund & Burgess, 1996). These models posit that semantic representations are created through the co-occurrences of words together within a body of text and suggest that words with similar meanings will appear together in similar contexts (Riordan & Jones, 2011). On the other hand, connectionist models of semantic memory (e.g., Rogers & McClelland, 2006; Rumelhart, McClelland, & Group, 1986) portray the semantic network 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).

Semantic overlap between concepts can measured in several ways. Feature production tasks (Buchanan, Holmes, Teasley, & Hutchison, 2013; McRae, Cree, Seidenberg, & McNorgan, 2005; Vinson & Vigliocco, 2008) provide one means of generating semantic word norms. 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. First, 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), with COS values ranging from 0 (completely 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 both 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 using COS values over JCN values is the limitation of JCN being tied to a somewhat static dictionary database, while a semantic feature production task can be used on any concept to calculate COS values. However, JCN values are less time consuming to obtain if both concepts are in the database (Buchanan et al., 2013).

Finally, indirect associations computed from a large dataset can also be used as a measure semantic overlap, and indeed may provide a better measure semantic relatedness relative to feature production norms. De Deyne et al. (2013b) constructed a semantic network based on the distributions of associations (e.g., indirect associates) by converting free association data taken from the SWOW project into a weighted semantic network. Computing the cosine overlap between the distribution of free association answers on any two concepts within this network provides a useful measure of meaning. Discussion of these measures then leads to the question of whether each one is truly assessing some unique concept or if they simply tap into various elements of our overall linguistic knowledge. Previous clustering and factor analyses by Maki and Buchanan (2008) indicates that there are potentially three separate latent structures represented by these various measures of similarity – associative, semantic, and thematic types of relatedness. However, another interpretation of their results is that the data collection of the measurement matters – variables that are based on participant responses to cued stimuli grouped together, while text-corpora based and WordNET based similarity measures separated into separate factors. By using the participant responses from SWOW to measure indirect association, we draw from a larger, newer set of data and resolve a potential confound of conflating measurement techniques.

Application to Judgment Studies

Traditional judgment of learning tasks (JOL) can be viewed as an application of the PAL paradigm; participants are given pairs of items and are asked to judge how accurately they would be able to correctly respond with the target with the cue on a recall task. Judgments are typically made out of 100, with a participant response of 100 indicating full confidence in recall ability. In their 2005 study, Koriat and Bjork examined overconfidence in JOLs by manipulating associative relations (forward strength from Nelson et al. (2004)) 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 (i.e., the likelihood of the target when shown the cue) 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. For example, the pair *bird* - *feather* in the SWOW norms appears to have a low forward strength (.031) and a higher backward strength (.199). However, the indirect relation between *bird* and *feather is .063. Therefore, it is important to investigate what may lead to the perceived relatedness between the item pairs and resulting in inflated JOLs.

The JOL task can then be manipulated to investigate perceptions of word pair relation 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 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 normed item pairs and found that responses were greatly overestimated relative to the actual normed overlap strength for pairs that were weak associates, while underestimated for 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 (JAM) function provides one means of visualizing the influence various cognitive biases have on associative memory judgments. By plotting the judged values against the word pair's normed associative strength, a fit line can be calculated which displays the calibration of JAM ratings relative to normed associative strength. When plotted, these judgments characteristically have a high intercept (an overestimation bias) along with a shallow slope (low sensitivity to changes in normed pair strength). Figure 1 illustrates this function. Overall, the JAM function has been shown to be highly reliable and generalizes 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, the Valentine and Buchanan (2013) study extended this research to include judgments of semantic memory with the same results. Finally, De Deyne et al. (2013a) found that JAM ratings for weak and moderate associates are best predicted by continuous response association norms relative to traditional free association norms.

The discrepancy between direct association strength and JAM ratings is noteworthy because on the surface, the two tasks should each be tapping into the same concept of associative overlap. One explanation for this provided by Maki (2007a) is that judgment tasks are more easily influenced by outside factors such as the availability heuristic. Thus, it may be that the act of viewing the cue-target pair together at the time of judgment interferes with individuals' ability to consider other target words that are related to the cue, thereby inflating (or reducing) the perceived relatedness between the items (Maki, 2007a). Indeed, work by (Nelson & Dunlosky, 1991) has shown this to be the case when eliciting judgments of learning, as JOLs made after a delay tend to be more accurate relative to those made immediately in the presence of the studied information. Further, the influence of indirect relations and their potential interaction on judgments have not been investigated.

The present study expanded upon previous JAM studies by examining recall and judgments for three types of judgments of relatedness (JORs) with the goal of exploring the underlying memory network that is used for each of these cognitive processes as described above. To date, no study has investigated the how three types of concept information affect these judgment and recall processes within the context of one unified study. As such, we tested four hypotheses, which were based upon previous research on JAM and semantic memory models.

First, we sought to expand upon previous Maki (2007b), Maki (2007a), Buchanan (2010), and Valentine and Buchanan (2013) research by including three types of JORs in one experiment, while replicating JAM bias and sensitivity findings. We used the SWOW norms calculating direct and indirect relations to best capture the continuum of similarity between concepts. These values were used to predict each type of JOR, and we calculated average slope and intercept values for each participant. 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. We also examined the frequency of each predictor being the strongest variable to predict an individual judgment condition. Thus, we are interested in exploring whether judgment findings replicate across judgment type and with new measurement variables available through SWOW (rather than each individually, as tested in previous JOL and JAM publications), which expands our knowledge on how the judgment process taps into the underlying memory network.

Next, we explored the predictions from semantic network models that the relation between associations and semantics would be bidirectional in nature (i.e., both types of knowledge interconnected in memory). Therefore, we expected to find an interaction between direct and indirect association norms when predicting JORs. We used multilevel modeling to examine the interaction of these norms in relation to participant judgments.

We then extended these analyses to include recall as the dependent variable of interest. We tested for the interaction of database norms in predicting recall by using a multilevel logistic regression, while controlling for judgment condition and rating. We expected to find that database norms would show differences in recall based on the levels of 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). Because judgment and recall are different cognitive processes, we used this hypothesis to examine how memory networks may be differently interactive for memory in comparison to judgment.

Finally, we examined if the judgment slopes from Hypothesis 1 would be predictive of recall. Whereas the recall model used to test our third hypothesis examined the direct relationship of word relatedness on recall, the goal of this analysis was to explore whether participant sensitivity to word relatedness was could also be used to predict recall. For this analysis, we used a multilevel logistic regression to control for multiple judgment slope conditions. This hypothesis combines both cognitive processes into one analysis, to explore how judgment ability (i.e., slopes) would impact the memory process.

Method

Participants

A power analysis was conducted using the *simR* package in *R* (Green & MacLeod, 2016). This package uses simulations to generate power estimates for mixed linear models created from the *lme4* package in *R* (Bates, Mächler, Bolker, & Walker, 2015). The results of this analyses suggested a minimum of 35 participants would be required to detect an effect. However, because power often tends to be underestimated, we extended participant recruitment as funding permitted. The data in this experiment was collected in two waves recruiting from Amazon's Mechanical Turk, which is a website that allows individuals to host projects and connects them with a large pool of respondents who complete them for small amounts of money (Buhrmester, Kwang, & Gosling, 2011). In the first wave, a total of 112 participants were recruited, and in the second wave, 221 participants were recruited. Participant responses were screened for a basic understanding of the study's instructions. Responses were rejected for participants who entered related words when numerical judgment responses were required, and for participants who responded to the cue words during the recall phase with sentences or phrases instead of individual words. Those that completed the study correctly were compensated \$1.00 for their participation in wave one, and \$2.00 for their participation in wave two. The second wave of participants was sponsored by graduate thesis funding provided by the Missouri State University Graduate College.

Materials

The stimuli used were 126 words pairs of varying relatedness which were created from the Buchanan et al. (2013) word norm database and website. These pairs were evenly split into sixty-three for wave one and wave two of the study. Pairs were originally selected by using forward strength (FSG; Nelson et al., 2004), semantic feature overlap cosine values (COS) from Buchanan et al. (2013), and Latent Semantic Analysis cosine values (Landauer & Dumais, 1997; Landauer, Foltz, Laham, Folt, & Laham, 1998) based on previous research on how word pair psycholinguistic variables overlap (Maki & Buchanan, 2008). The selected stimuli included a range of values for each variable. Table 1 displays 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. Pairs with low, medium, and high FSG and LSA were then selected, when avaliable. Given the measurement questions raised in the introduction, the direct association from the SWOW norms will be used as the measure of first order association. Given De Deyne et al. (2013a)'s work on continuous association, the response set from all three responses were used. The direct association provided in these norms is calculated as the number of participants who provided the target to the cue divided by the number of possible answers (i.e., participants * responses). This calculation, therefore, has an upper limit of approximately ~33%, even if every participant listed a target word to a cue. The JOR task assumes the range of direct association is 0 to 100 (or 0-1 proportion), and the SWOW direct association (DA) was normalized using:

$$\frac{DA - Min(DA)}{Max(DA) - Min(DA)}$$

Indirect association (IA) was calculated by comparing the distribution of responses for each concept. Therefore, if the concepts were *bird* and *feather*, the two association sets were combined and the cosine between the response frequencies was calculated. Cosine indicates a measure of overlap in the response distributions, where 0 indicates no overlapping responses, while 1 indicates perfectly overlapping response frequencies (see Buchanan, Valentine, & Maxwell, 2019 for more on cosine feature overlap). DA and IA averages are provided in 1. The study was built online using Qualtrics, and three surveys were created to counter-balance the order in which judgment conditions 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 phase, JORs were elicited by presenting participants with word pairs and asking them 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 described 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*. In each judgment, participants were then given three concepts (*lost*, *old*, *article*) and were asked to come up with words that they felt were related to that type of relation.

After viewing the examples at the start of the block, participants completed the JOR task. Each block contained a set of instructions which were contingent upon the type of JOR being elicited. For example, instructions in the associative block asked participants to estimate how many individuals out of 100 they expect would respond to the cue word with a given target, instructions for semantic JORs asked participants to indicate the percent of features shared between two concepts, and instructions for the thematic JOR task asked participants to base ratings on how likely to words would be used together in the same story. The complete experiment can be found at http://osf.io/y8h7v, which contains the exact instructions given to participants for each block and displays the structure of the study. All instructions were modeled after Buchanan (2010) and Valentine and Buchanan (2013).

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

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

Results

Data Processing and Descriptive Statistics

First, the results from the recall phase of the study were coded as zero for incorrect responses, one for correct responses, and NA for participants who did not complete the recall section (all or nearly all responses were blank). All word responses to judgment items were deleted and set to missing data. The final dataset was created by splitting the initial data file into six sections (one for each of the three experimental blocks and their corresponding recall scores). Each section was individually melted using the *reshape* package in *R* (Wickham, 2007) and was written as a csv file. The six output files were then combined to form the final dataset. Code is available on our OSF page embedded inline with the manuscript in an *R* markdown document written with the *papaja* package (Aust & Barth, 2017). With 333 participants, the dataset in long format included 20979 rows of potential data (i.e., 333 participants * 63 JORs). 15 out of range JOR data points (> 100) were corrected to NA. Missing data for JORs or recall were then excluded from the analysis, which included word responses to judgment items (i.e., responding with *cat* instead of a number when prompted to provide a JOR). These items usually excluded a participant from receiving Amazon Mechanical Turk payment, but were included in the datasets found online. In total, data points2266 were excluded (679 JOR only, 1019 recall only, 568 both), leading to a final N of 18713 observations. Recall and JOR values were then screened for outliers using Mahalanobis distance at $p < .001$, and no outliers were found (Tabachnick & Fidell, 2012). To screen for multicollinearity, we examined correlations between judgment items, DA, and IA. All correlations were rs $< .26$.

The mean JOR for the associative condition ($M = 59.40$, $SD = 29.52$) was lower than the semantic ($M = 64.15$, $SD = 29.74$) and thematic ($M = 69.50$, $SD = 28.20$) conditions. A multilevel model was examined to determine if these JOR values were significantly different using participants as a random 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, which makes these models advantageous for multiway repeated measures designs (Gelman, 2006). Associative judgments were lower than both semantic ($t(19407) = 10.40$, $p < .001$), and thematic judgments ($t(19407) = 22.25$, $p < .001$). Semantic judgments in turn were lower than thematic judgments ($t(19407) = 11.85$, $p < .001$).

Recall averaged around 60% for all three conditions: associative $M = 5903.95$, $SD = 4917.99$; semantic $M = 6256.76$, $SD = 4839.85$; thematic $M = 6011.76$, $SD = 4896.94$. A separate multilevel model indicated that associative recall was lower than semantic recall ($t(19064) = 4.63$, $p < .001$), but not thematic recall ($t(19064) = 1.37$, $p = .169$). Semantic recall scores were higher than thematic recall scores ($t(19064) = -3.25$, $p = .001$).

JAM Slope Bias and Sensitivity

[1] "The observed noncentrality parameter of the noncentral t-distribution has exceeded 37.62 in magnitude (R's limitation for accurate probabilities from the noncentral t-distribution) in the function's iterative search for the appropriate value(s). The results may be fine, but they might be inaccurate; use caution." [1] "The observed noncentrality parameter of the noncentral t-distribution has exceeded 37.62 in magnitude (R's limitation for accurate probabilities from the noncentral t-distribution) in the function's iterative search for the appropriate value(s). The results may be fine, but they might be inaccurate; use caution." [1] "The observed noncentrality parameter of the noncentral t-distribution has exceeded 37.62 in magnitude (R's limitation for accurate probabilities from the noncentral t-distribution) in the function's iterative search for the appropriate value(s). The results may be fine, but they might be inaccurate; use caution."

First, we sought to replicate bias and sensitivity findings from previous research while expanding the JAM function to include judgments based on three types of memory. DA and IA were used to predict each type of relatedness judgment. JOR values were divided by 100, so as to place them on the same scale as the direct and indirect association. Slopes and intercepts were then calculated for each participant's ratings for each of the three JOR conditions, as long as they contained at least nine data points out of the twenty-one that were possible. Single sample *t*-tests were then conducted to test if slope and intercept values significantly differed from zero. See Table 2 for means and standard deviations. Slopes were then compared to the JAM function, which is characterized by high intercepts (between 40 and 60 on a 100 point scale) and shallow slopes (between 20 and 40). Because of the scaling of our data, to replicate this function, we should expect to find intercepts ranging from .40 to .60 and slopes in the range of 0.20. to 0.40. Intercepts for associative, semantic, and thematic JORs were each significant, and all fell within or near the expected range. Overall, thematic JORs had the highest intercept at .606, while JORs elicited in the semantic condition had the lowest intercept at .518.

The JAM slope was successfully replicated for DA in all three conditions with slopes falling in the expected range of 0.20 to 0.40. For associative judgments, the indirect relation - which is thought to be representative of semantic relatedness - did not predict judgments. In the thematic judgment condition, the indirect values were positive, indicating contribution of both direct and indirect values to the judgments, which were described as a mix of both relation types. Last, the semantic judgment condition showed that both direct and indirect relations were important with the highest indirect contribution of the three judgment types, indicating differences in focus of judgment tap different relations to meet task demands. Overall, JAM slopes were successfully replicated in each JOR condition, the high intercepts and shallow slopes present across conditions were indicative of overconfidence and insensitivity in participant JORs.

Additionally, we examined the frequency that each predictor variable was the strongest predictor for each of the three JOR conditions. For the associative condition, the direct association was the strongest predictor for 67.3% of the participants. This distinction was less pronounced when examining the semantic and thematic JOR conditions. In the semantic condition, DA was 52.1% of participants, and in the thematic condition, DA was 54.1% of participants. These results mirror the slope values, such that direct association is strongest when asked to judge associative relations, while a more even split between direct and indirect predictors was found when asked to consider semantic and thematic relations.

Interaction between Relation when Predicting Judgments of Relatedness

The goal of next analysis was to test for an interaction between direct and indirect association when predicting participant JORs. First, the database norms were mean centered aide in interpretation. 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 DA and IA 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. This analysis resulted in a significant interaction between DA and IA (β = -0.19, $p < .001$), which is examined below in a simple slopes analysis. Table 3 includes values for main effects, two-way interaction, and the simple slopes.

To investigate this interaction, simple slopes were calculated for low, average, and high levels of indirect association. This variable was chosen for to show the effects of direct associations across levels of indirect association. At low levels of indirect relation, and thus low levels of the semantic network, we found the largest β for direct association, 0.33. As indirect relation increases, we found decreasing predictiveness of direct relation, average direct $\beta = 0.29$, and high direct $\beta = 0.25$. Figure 2 displays the two-way interaction with this seesaw type effect, indicating that higher semantic network relation results in lower usefulness of direct associative relation.

Interaction between Relation when Predicting Recall

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 (Bates et al., 2015), testing the interaction between DA and IA when predicting participant recall. As with the previous hypothesis, we controlled for JOR condition and, additionally, covaried JOR ratings. Participants were used as a random intercept factor. Judged values were not a significant predictor of recall, (β = 0.04, p .512). A significant interaction was detected between direct and indirect relations (β = -1.30, p .008). See Table 4 for main effects, interaction, and simple slopes.

The same moderation process used in Hypothesis 2 was then repeated, with simple slopes calculated at low, average, and high levels of indirect association. The same pattern of results emerged where low levels of indirect association resulted in the largest contribution of direct association $\beta = 1.90$. As indirect association increased, direct association coefficients decreased, average direct $\beta = 1.63$, and high direct $\beta = 1.37$. Thus, the cognitive processes of recall and judgment appear to operate similiarly on the memory network.

Predicting Recall with JAM Slopes

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

Buchanan, E. M., Valentine, K. D., & Maxwell, N. P. (2019). English semantic feature production norms: An extended database of 4436 concepts. *Behavior Research Methods*. doi:10.3758/s13428-019-01243-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

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. (2013a). Associative strength and semantic activation in the mental lexicon: evidence from continued word associations. In *Proceedings of the 35th annual conference of the cognitive science society* (pp. 2142–2147).

De Deyne, S., Navarro, D. J., & Storms, G. (2013b). Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior Research Methods*, *45*(2), 480–498. doi:10.3758/s13428-012-0260-7

Deese, J. (1965). *The structure of association in language and thought.* Baltimore, MD: The Johns Hopkins University Press.

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

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

Huff, M. J., & Hutchison, K. A. (2011). The effects of mediated word lists on false recall and recognition. *Memory & Cognition*, *39*(6), 941–953. doi:10.3758/s13421-011-0077-0

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

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

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

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

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/BF03192726

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

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely acurate at predicting subsequent recall: The delayed-JOL effect. *Psychological Science*, *2*(4), 267–270. doi:10.1111/j.1467-9280.1991.tb00147.x

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., & Team, R. C. (2017). nlme: Linear and nonlinear mixed effects models. Retrieved from https://cran.r-project.org/package=nlme

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

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 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

	Semantic Feature Overlap COS		
Variable	Low	Average	High
Semantic Feature Overlap COS	.09 (.10)	.45 (.09)	.75 (.05)
Forward Strength FSG	.08 (.11)	.15 (.17)	.19 (.24)
Latent Semantic Analysis LSA	.25 (.17)	.39 (.18)	.47 (.18)
Direct Association DA	.12 (.16)	.23 (.23)	.27 (.29)
Direct Association IA	.10 (.14)	.25 (.17)	.39 (.18)

Note. Standard deviation values are in parentheses.

Table 2

Summary Statistics for Hypothesis 1 t-Tests

Variable	M	SD	t	df	p	d	$95\% CI$
Associative Intercept	.522	.201	45.905	311	$<$ .001	2.599	2.364 - 2.829
Associative DA	.326	.299	19.269	311	$<$ .001	1.091	0.950 - 1.231
Associative IA	.027	.293	1.624	311	.105	0.092	-0.019 - 0.203
Semantic Intercept	.518	.205	44.635	312	$<$ .001	2.523	2.294 - 2.748
Semantic DA	.315	.301	18.469	312	$<$ .001	1.044	0.906 - 1.181
Semantic IA	.243	.324	13.305	312	$<$ .001	0.752	0.626 - 0.877
Thematic Intercept	.606	.182	59.245	315	$<$ .001	3.333	3.048 - 3.613
Thematic DA	.277	.266	18.536	315	$<$ .001	1.043	0.905 - 1.179
Thematic IA	.137	.282	8.653	315	$<$ .001	0.487	0.370 - 0.603

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

Table 3

MLM Statistics for Hypothesis 2

Statistic	Coefficient	SE	t	p
Intercept	0.59	0.01	77.47	$< .001$
Semantic Judgment	0.05	0.00	10.73	$< .001$
Thematic Judgment	0.10	0.00	23.11	$< .001$
Z Direct	0.29	0.01	38.98	$< .001$
Z Indirect	0.14	0.01	15.92	$< .001$
Z Interaction	-0.19	0.05	-3.86	$< .001$
Z Direct Low	0.33	0.01	26.40	$< .001$
Z Direct High	0.25	0.01	20.58	$< .001$

Note. Direct and indirect association were mean centered. The table shows main effects and interactions predicting participant judgments. df = 19404

Table 4

MLM Statistics for Hypothesis 3

Statistic	Coefficient	*SE*	*Z*	*p*
Intercept	0.42	0.08	5.09	$< .001$
Semantic Judgment	0.18	0.04	4.38	$< .001$
Thematic Judgment	0.05	0.04	1.10	.270
Judged Value	0.04	0.07	0.66	.512
Z Direct	1.63	0.08	20.32	$< .001$
Z Indirect	0.32	0.09	3.71	$< .001$
Z Interaction	-1.30	0.49	-2.67	.008
Z Direct Low	1.90	0.13	14.58	$< .001$
Z Direct High	1.37	0.12	11.09	$< .001$

Note. Direct and indirect association were mean centered. The table shows main effects and interactions predicting recall.

Table 5

MLM Statistics for Hypothesis 4

Variable	b	SE	z	p
(Intercept)	-0.108	0.270	-0.399	.690
ACOS	-0.098	0.232	-0.423	.673
AFSG	1.168	0.257	4.543	$<$ .001
AIntercept	0.487	0.392	1.241	.214
(Intercept)	-1.159	0.313	-3.709	$<$ .001
SCOS	1.266	0.230	5.497	$<$ .001
SFSG	1.313	0.253	5.185	$<$ .001
SIntercept	1.799	0.428	4.199	$<$ .001
(Intercept)	-1.004	0.319	-3.149	.002
TCOS	1.054	0.253	4.170	$<$ .001
TFSG	1.089	0.265	4.109	$<$ .001
TIntercept	1.738	0.424	4.094	$<$ .001

Note. Each judgment-database bias and sensitivity predicting recall for corresponding judgment block. A: Associative, S: Semantic, T: Thematic.

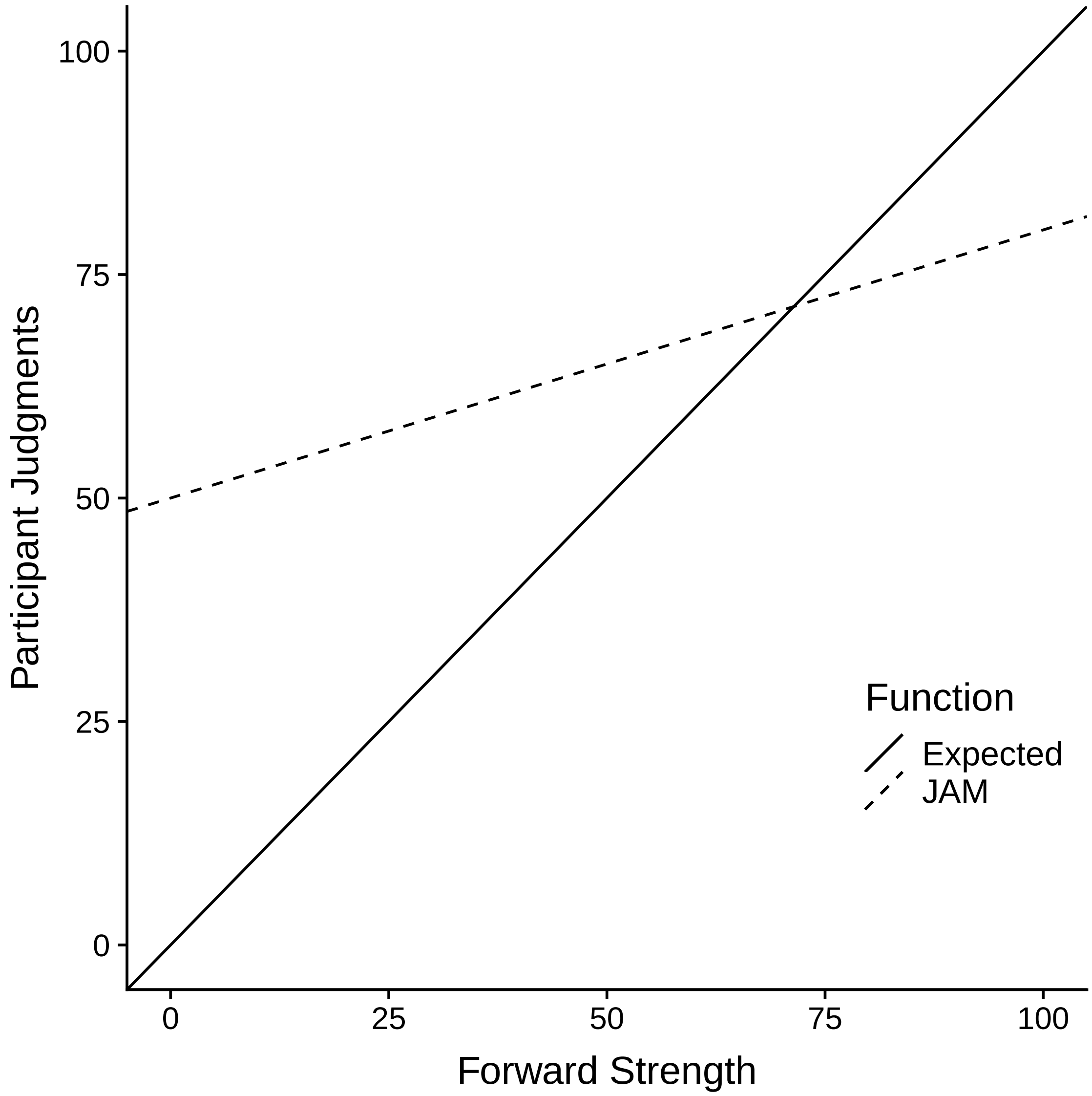

Figure 1. JAM slope findings from Maki (2007a). JAM is characterized by a high intercept (between 40 and 60) and a shallow slope (between 0.20 and 0.40). The solid line shows expected results if judgment ratings are perfectly calibrated with association norms.

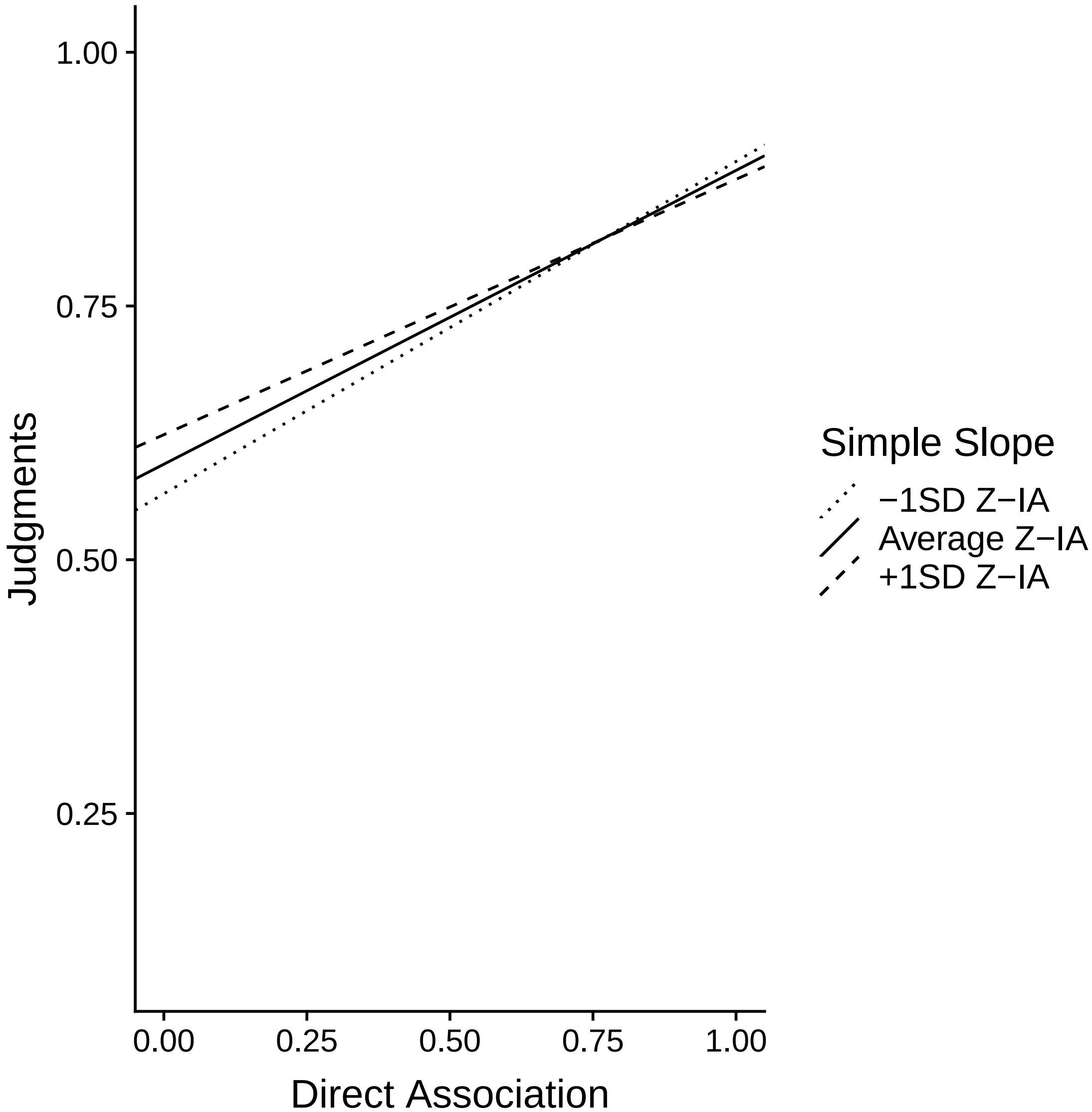

Figure 2. Simple slopes graph displaying the slope of direct association when predicting JORs at low, average, and high indirect association. All variables were mean centered.

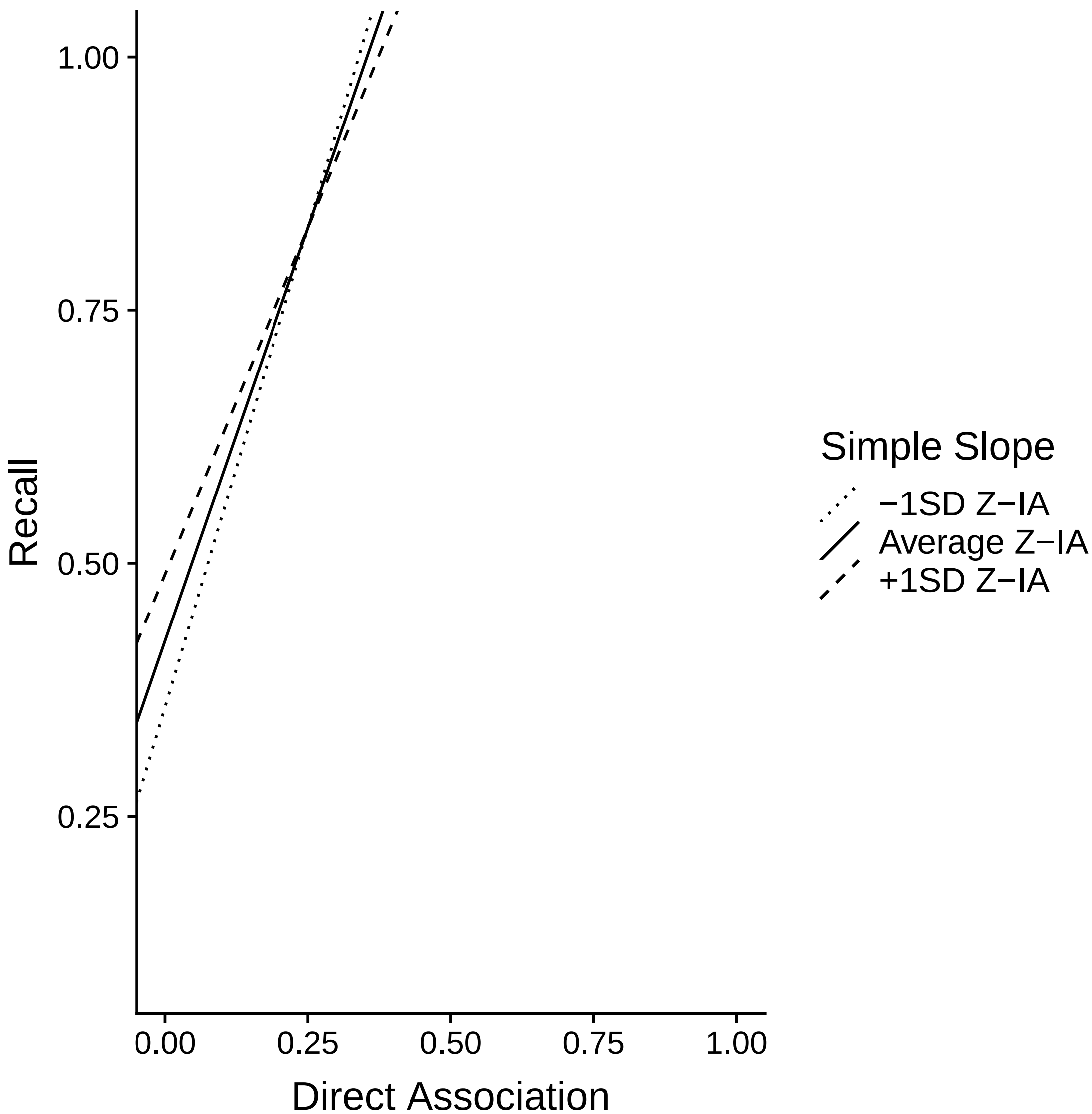

Figure 3. Simple slopes graph displaying the slope of direct association when predicting recall at low, average, and high indirect association. All variables were mean centered.