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, @Smythe1968 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 [@Richardson1998], with modern studies shifting their focus towards a broad range of applied topics such as how PAL is affected by aging [@Hertzog2002], its impacts on second language acquisition [@Chow2014], and even in evolutionary psychology [@Schwartz2013].

Early PAL studies routinely relied on stimuli generated from word lists that focused extensively on measures of word frequency, concreteness, meaningfulness, and imagery [@Paivio1969]. 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 [@Buchanan2010], 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 [@﻿Nelson2000].

## 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 @DeDeyne2013 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 [@Nelson2000]. 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 [@Nelson2000]. Thus, the free association process can be thought of as generating an index that contains the relative accessibility of related words in memory [@Nelson2004].

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; @Nelson2004]. 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, @DeDeyne2013; @DeDeyne2018] 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 [@DeDeyne2013].

## 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 [@DeDeyne2013; @Deese1965]. 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 @Huff2011 for a review of mediated associates] and is one of the underlying factors behind distributional models of semantic memory [e.g., Latent Semantic Analysis, @Landauer199; Hyperspace Analogue to Language Model, @Lund1996]. 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 [@Riordan2011]. On the other hand, connectionist models of semantic memory [e.g., @Rogers2006; @Rumelhart1986] 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 [@Jones2015].

Semantic overlap between concepts can measured in several ways. Feature production tasks [@McRae2005; @Vinson2008; @Buchanan2013] 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 [@McRae2005]. 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 [@Maki2004], 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 [@Miller1995]. The JCN value is then computed by summing together the difference of each concept and its lowest super-ordinate [@Jiang1997; @Maki2004]. 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 [@Buchanan2013].

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. @DeDeyne 2013 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 @Maki2008 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 @Nelson2004) 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. Take for example the pair \*feather\* - \*bird\*, which has an FSG of .051 based on the USF norms, yet also has a cosine value of .272 (suggesting low to moderate feature overlap). As such, some of the overconfidence in JOLs may be attributed to more than just item associations, as paired items may also be connected by similar themes or share certain features, increasing the perceived relatedness between the item pairs and resulting in inflated JOLs.

The traditional judgment of learning task (JOL) can then be manipulated to investigate perceptions of word pair relationships by having participants judge how related they believe the cue and target items to be [@Maki2007a; @Maki2007]. 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 [@Maki2007], which mimics how association word norms are created through free association tasks. @Maki2007a 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 @Koriat2005 findings for relatedness judgments based upon associative memory, rather than judgments based on learning.

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 [@Nelson1991] 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.

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 \@ref(fig:makislope) 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 [@Maki2007]. 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 [@Valentine2013]. Additionally, the @Valentine2013 study extended this research to include judgments of semantic memory with the same results. Finally, De Deyne et al. (2013b) found that JAM ratings for weak and moderate associates are best predicted by continuous response association norms relative to traditional free association norms.

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 @Maki2007, @Maki2007a, @Buchanan2010, and @Valentine2013 research by including three types of JORs in one experiment, while replicating JAM bias and sensitivity findings. We used database norms measuring associative and semantics overlap to predict each type of JOR and 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 its own judgment condition (i.e., how often association was the strongest predictor of associative JORs, etc.). Thus, we are interested in exploring whether judgment findings replicate across a range of variables and covariates (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.