**Intro**

The study of cognition has a rich history of exploring the role of association in human memory. One key finding is that elements of cognitive processing play a critical role in how well an individual retains learned information. Throughout the mid-20th century, much research was conducted that 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. 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 effected by aging (Hertzog et al. 2002), its effects on second language acquisition (Chow 2014), and even in evolutionary psychology (Schwartz and 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, 1965). 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). Additionally, these variables capture psycholinguistic measurement 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 relation in a reliable manner. As a result, free association norms have become a common means of indexing associative strength between word pairs (Nelson McEvoy Schreiber reference). 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 et al. 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 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). Using this technique, researchers have developed databases of associative word norms that can be used to generate stimuli with a high degree of reliability. Many of these databases are now readily available online, with the largest one consisting of over 72,000 associates generated from more than 5,000 cue words (Nelson et al. 2004).

Similar to association norms, semantic word norms provide researchers with another means of constructing stimuli for recall 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, Rumelhart et al. 1986, Rogers and McClelland 2006) 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 et al. 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 context (Riordan and Jones 2011).

Feature production tasks are a common means of producing semantic word norms (e.g., McRae et al. 2005, Vinson and Vigliocco 2008, Buchanan et al. 2014). 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 (COS) between them (Maki et al. 2004). For example, the pair HORNET – WASP has a COS of .88, indicating high overlap between the two concepts. Feature overlap can also be measured by JCN, which involves calculating the information content for each concept and the lowest superordinate shared by each concept using an online dictionary, WordNET (cite). The JCN value is then computed by summing together the difference of each concept from their lowest superordinate (Jiang and Conrath 1997, Maki et al. 2004). The advantage to using COS values over JCN values is the limitation of JCN tied to a somewhat static dictionary database, as 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. 2014).

Semantic relations can be broadly described as being taxonomic or thematic in nature. Whereas taxonomic relationships focus on the connections between features and concepts within categories (e.g., BIRD – PIDGEON), thematic relationships center around the links between concepts and an overarching theme or scenario (e.g., BIRD – NEST, Jones and Golonka 2012). Jouravlev and McRae (2015) provide a list of 100 thematic relatedness production norms, which were generated through a task similar to feature production in which participants were presented with a concept and were asked to list names of other concepts they believed to be related. Distributional models of semantic memory also lend themselves well to the study of thematic word relations. Because these models are text based and score word pair relations in regard to their overall context within a document, they assess thematic knowledge as well as semantic knowledge. Additionally, text based models such as latent semantic analysis (LSA) are able to account for both the effects of context and similarity of meaning, bridging the gap between associations and semantics (Landauer et al 1998).

Discussion of these measures naturally raises the question of whether they truly assess unique concepts or simply tap into our overall linguistic knowledge. Taken at face value, word pair associations and semantics word relations appear to be vastly different, yet the line between semantics/associations and thematics is much more blurred. While thematic word relations are indeed an aspect of semantic memory and includes word co-occurrence as an integral part of creation, themes appear to be indicative of a separate area of linguistic processing. Previous research by Maki and Buchanan (2008) appears to confirm this theory. Using clustering and factor analysis techniques, they analyzed multiple associative, semantic, and text based measures of associative and semantic knowledge. Their findings suggest 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, and even the first studies on semantic priming used association word norms for stimuli creation (Meyer & Schvaneveldt, 1971; Meyer, Schvaneveldt, & Ruddy, 1975; Lucas, 2000). 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).

**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 match the target with the cue on a recall task. Judgments are typically made out of 100, with 100 indicating full confidence recall ability. In their 2005 study, Koriat and Bjork examined overconfidence in 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. Koriat and Bjork proposed that this overconfidence was the product of a foresight bias, which they considered an inverse of the widely investigated hindsight bias.

JOL tasks can be manipulated to investigate perceptions of word pair relationships by having participants judge how related they believe the stimuli to be (Maki, 2007 or 08? Put both the JAM citations here). Judged values 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 references since he created it), which mimics how the association word norms were created. Maki (2007a) investigated such judgments within the context of associative memory and found that responses greatly overestimated the strength of relationship for pairs that were weak associates, while underestimating strong associates; thus replicating the Koriat and Bjork (2005) findings for judgments on memory, rather than 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). The JAM function was found to be highly reliable and generalized across multiple variations of the study, with item characteristics such as word frequency, cue set size (QSS), and semantic similarity having a minimal influence on it (Maki 2007b). 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 judgments of semantic memory with the same results.

The present study combined PAL and JAM to examine item recall within the context of judgment, while extending the JAM task to include judgments of semantic and thematic memory. Relationship strengths between word pairs were manipulated across each of the three types of memory investigated using previous research on databases to assure a range of relatedness. We tested the following hypotheses:

1. First, we sought to expand previous Maki (2007/8?), Buchanan (2009) and Valentine and Buchanan (2013) research to include three types of judgments of memory in one experiment, while replicating bias and sensitivity findings. We used the three database norms for association, semantics, and thematics to predict each type of judgment and calculated average slope and intercept values for each participant. We expected to find slope and intercept values that were significantly different from zero, as well as within the range of previous findings. Additionally, we examined the frequency of each predictor being the strongest variable to predict its own judgment condition (i.e. how often association was the strongest predictor of associative judgments, etc.).
2. Given the overlap in these variables, we expected to find an interaction between database norms in predicting participant judgments, controlling for judgment type. We used multilevel modeling to examine that interaction of database norms for association, semantics, and thematics in related to participant judgments.
3. These analyses were then extended to recall as the dependent variable of interest. We examined the interaction of database norms in predicting recall by using a multilevel logistic regression, while controlling for judgment type and rating. We expected to find that database norms would show differences in recall based on the levels other variables (the interaction would be significant), and that ratings would also positively predict recall (i.e. words that participants thought were more related would be remembered better).
4. Finally, we examined in the judgment slopes from Hypothesis 1 would be predictive of recall. Hypothesis 3 examined the direct relationship of word relatedness on recall, while this hypothesis explored if participant sensitivity to word relatedness was a predictor of recall. For this analysis, we used a multilevel logistic regression to control for multiple judgment slope conditions.