Running head: judgments of associative memory

From bad to worse: variations in judgments of associative memory

Erin Buchanan {1}

{1} Department of Psychology, University of Mississippi

Acknowledgments

Erin Buchanan

University of Mississippi

Department of Psychology

PO BOX 1848

University, MS 38677

Phone: 662-915-1075

Email: embuchan@olemiss.edu

Abstract

Four different groups were tested in variations of the Judgments of Memory task created by Maki (2007a). Participants judged word pairs on the strength of their relationship, performed a free association task (Nelson, McEvoy, & Schreiber, 2004), and judged how many words were related to a given cue word. As seen with Maki (2007a; 2007b), participants cannot discriminate between word pairs with high and low relationships. This finding was extended to show that people also cannot free associate word pairs with different numbers of associates, or even judge how many associates a cue word is linked to in memory. However, participants' judgment ability varied widely in each one of the tasks. Working memory and fluid intelligence were measured in order to examine metacognitive errors, but had little to no influence on participant ability to judge word pair relationships.

Keywords: associations, judgments of memory

From Bad to Worse: Variations in Judgments of Associative Memory

Are there more words that begin with the letter "K" or are there more words that have the letter "K" in the third position as in cake? When presented with this question, people generally have a difficult time deciding between "K" in the first position and "K" in the third position, but usually choose the first position (Kahneman & Tversky, 1973). K appears in the 3rd position in more than 3 times the number of words than the first position. This example portrays the complexity of making a judgment that involves searching through memory, which is examined in the judgments of learning (JOL) and judgments of memory literature (Koriat & Bjork, 2005). In a JOL task, participants are required to learn a set of materials and then make a confidence rating on how well they know the information. After they make these confidence ratings, participants are tested on the material. Confidence ratings of learned materials rarely match actual test performance and are generally higher than performance (Koriat, 2008).

A judgment of associative memory (JAM) task also requires access into memory, but not for recently learned items. Participants are shown a related word pair (LOST-FOUND) and asked to estimate the number of people who would give a target word (FOUND) if presented a cue word (LOST; Maki 2007a). Participants consistently overrate the relationship between associatively related words. People are also insensitive the strength of the relationship, which is seen in the JAM function (Maki, 2007b). Not only are the judgments biased upwards, but the slope of the JAM function is nearly flat, meaning participants cannot distinguish between lowly and highly related word pairs.

On the surface, the JAM task does not seem like the same metacognitive process as the JOL task because of the differences in procedure. The paired-associates task for JOLs has been

used since Arbuckle and Cuddy (1969), which is similar to the JAM paradigm. Participants are shown word pairs that have strong or weak relationships and asked to judge how well they will remember them. The strength of the relationship between word pairs predicted later performance, even if JOLs did not. Koriat and Bjork (2006) recently have shown that both forward strength of an associative word pair relationship (e.g. LOST to FOUND) and the backwards relationship (e.g. FOUND to LOST; see Nelson et al., 2004) have an effect on JOLs and recall. The perceived relationship between two words also influences the ability to accurately judge material learning (Koriat & Bjork, 2005).

The current research is concerned with the ability to make judgments of memory on associatively related words. The judgment of learning literature sheds light into possible explanations of why JAM is biased. Mirroring Koriat and Bjork (2006), Maki (2007b) has shown that the slope in the JAM function cannot be changed with training. He showed several groups of participants associatively related word pairs with feedback on the accuracy of their judgment estimates. During this training phase, some participants continued to get the same word pairs and others got completely new pairs. The group of participants who received the same pairs continued to improve because of rote learning, while the training group continued to have biased judgment estimates. Then each group was tested on a completely different set of word pairs. In this test, the intercept of JAM had decreased to lower the overall bias, but the slope had not changed from previous findings.

One problem with making associative judgments is that words often have many links in memory (e.g. COMPUTER is associatively related to over 30 words). The inability to judge the relationship between COMPUTER and MOUSE may be due to the other competing cues in memory like PROGRAM. In the associative database, the probability of a cue and all its target

words has an upper limit of 1, meaning that if all the words related to a cue were presented, the total JAM score would total to 100 percent. In a second experiment, Maki (2007b) asked participants to rate four target words for their relationship with one cue word. Their judgments were constrained so that they could only total 100 percent and would hopefully change both the intercept and the slope. Even when given the competing targets in memory, participants were still insensitive to the difference between strongly and weakly related word pairs.

The same effect is not seen when examining judgments of semantic memory, however. Maki, Krimsky, and Munoz (2006) had participants rate pairs of semantically related words drawn from the McRae, Cree, Seidenberg, and McNorgan (2005) semantic feature production norms and Maki, McKinley, and Thompson (2004) semantic distance norms. Word pairs were judged on the number of overlapping features and were quite accurate with semantic database norms. The correlation between word pair judgments and relationships was between .5 and .9, while associative judgments tend to had much shallower slopes at around .2 to .3 (Maki, 2007a).

These two tasks simply have different instructions, but show dichotomous results when asked to judge information from memory. A look at the data from the previous associative judgments research shows that there is great variation in the slopes across participants (and was found in the current research). Variation in people's ability has often been described as the differences in working memory. Beginning with Daneman and Carpenter (1980), working memory has been related a wide array of cognitive tasks (Engle, Tuholski, Laughlin, & Conway, 1999). Working memory has been described as the active use of information in memory and is directly related to ability to manipulate direction of attention (Kane, Conway, Hambrick, & Engle, 2007).

Working memory would appear to explain the differences in judgment ability across participants because it would predict their ability to manipulate the information pulled from memory. Logically, higher working memory spans would be able to hold and use more information than lower memory spans. This advantage should improve judgments because of the ability to consider more information while formulating a judgment. Therefore, the automated version of the operation span task (AOSPAN; Unsworth, Heitz, Schrock, & Engle, 2005) was used to examine the relationship between working memory and the JAM task.

This project also examined the nature of the JAM task that seems to separate it from other judgment tasks, such as the semantic judgment task. Word pairs for the JAM task are selected from the Nelson et al. database of free association norms. These norms were created by asking participants to name the first word that came to mind after being presented with a cue word. Word pair strength is the probability of one target word given a cue word, which is the estimated judgment in a JAM task. The JAM task, therefore, is the inverse of how the associative norms are created.

Judgments could possibly be improved by asking participants to perform a task that mirrors the free association task rather than the estimation of other people's responses. Several different versions of this task were created. In one group, participants were told free associate (i.e. name as many words as possible) to a list of given cue words. Another group was asked to estimate how many associates a set of cue words was related to and similarly a group was asked to simply guess if the cue word had many associates or only a few associates. A final group was asked to perform the traditional JAM task. All of the groups received instructions that explained the nature of associative memory, a free association task and feedback examples on how to perform the task.

Method

Participants.

A total of 133 participants were tested from Texas Tech University and the University of Mississippi. Group A contained 44 participants, and they were selected from the human subjects' pool from Texas Tech University. Group B tested 25 participants, Group C tested 26 participants, Group D tested 38 participants and were all selected from the human subjects' pool at the University of Mississippi.

Apparatus.

The participants were tested individually using personal computers equipped with 15-in. color monitors.

Materials.

Several measures were collected to test the components of judgments of associative memory: the AOSPAN (Unsworth et al., 2005), the Raven Progressive Matrices (Raven, Raven, & Court, 1998), and the Judgments of Associative Memory (JAM) task (Maki, 2007a). The AOSPAN is an automated version of Turner and Engle's (1989) operation span task, which measures working memory. Participants are first required to solve a math problem (IS (8/2) − 1 = 1?), followed by a letter to remember. These problem-letter combinations were presented in sets of 3 to 7, totaling 75 problems overall. After a set of problems, participants recall the letters presented in the order they were presented. The overall working memory score is the number of correct letters in their correct presentation order. The Raven Progressive Matrices is a measure of reasoning and fluid intelligence. Participants are shown a 3X3 matrix of 8 pieces that consist

of a pattern going both across and down. They are required to fill in the pattern with a 9th piece from a selection of 8 options. There are 36 problems total, and their score is the total number of correct patterns completed in 25 minutes.

In the JAM task, the nature of associative memory was described first (see Maki, 2007 for complete instructions). Group A then completed the traditional JAM task by rating related word pairs on their associative strength, which were selected from the Nelson et al. (2004) database. These word pairs were used by Maki (2007a) and have a range of associative values from very low (forward strength FSG = .1) to medium (FSG = .4) to very high (FSG = .7). They were told to name "how many college students out of 100 would write the SECOND cue word to the FIRST target word". They typed their responses in a box next to each word pair.

Group B, C, and D received the same basic instructions on how word pairs are associatively related. Participants were then asked to do a practice free association task. They were given a target word (LOST) and asked to name words they thought were related to the target (FOUND). Next, several examples of related word pairs were shown to help explain that some words have only a few associates (3-5 targets, i.e. MELODY), and other words have many associates (20-30 targets, i.e. COMPUTER). Participants were then given a list of 30 words to fill in as a free association task. These 30 words were selected from the Nelson et al. database so that 15 of the words had a small number of associates (3-5) and 15 of the words had a large number of associates (20-30). Group B was told to fill as many related words they could think of while remembering that the words had different numbers of associates. Group C was told to list the number of words they thought were related to the given target word. Group D simply had to do a discrimination task, where they labeled which words had a high number of associates and which words had a low number of associates.

Procedure.

Participants were first tested on the AOSPAN task and following the completion of this task, participants were given a chance to take a five-minute break. The Raven Progressive Matrices was performed second, again followed by a five-minute break. Finally, the Judgments of Associative Memory task was last for each group. The only variation between groups was the type of judgment task given. After the Judgments of Associative Memory task, participants were debriefed and given course credit.

Results

For each participant, a regression slope was calculated to examine the relationship between judgment ratings and database scores. The beta in the regression equation would be equal to one if participants were perfectly able to judge the difference between highly associated pairs and lowly associated pairs. For Group A, the forward strength variable (FSG) from the Nelson et al. (2004) database norms were used to predict each participant's judgment scores. For Groups B, C, D, the number of associates of each cue word (cue frequency) were used to predict participant's number of listed associates (Group B), predicted number of associates (Group C), and high/low values (Group D). For Group D, associates were categorized as high or low to match the type of judgment made by participants for the slope calculation.

The first analysis used individual difference measures to explain the variation in judgment ability within a group. Four multiple linear regressions were run using AOPSAN and Raven's scores predicting participant slopes for each group. AOSPAN and Raven's scores did not significantly predict variation in the participant slopes for a traditional judgments of memory task (Group A, $F(2,22) < 1$, $p = .844$, $R^2 = .015$), ability to free associate (Group B, $F(2,23) < 1$,

$p = .527$, $R^2 = .054$), ability to guess the number of associates linked to a cue word (Group C,

$F(2,35) = 1.121$, $p = .337$, $R^2 = .060$), or ability to judge if a cue word is linked to many

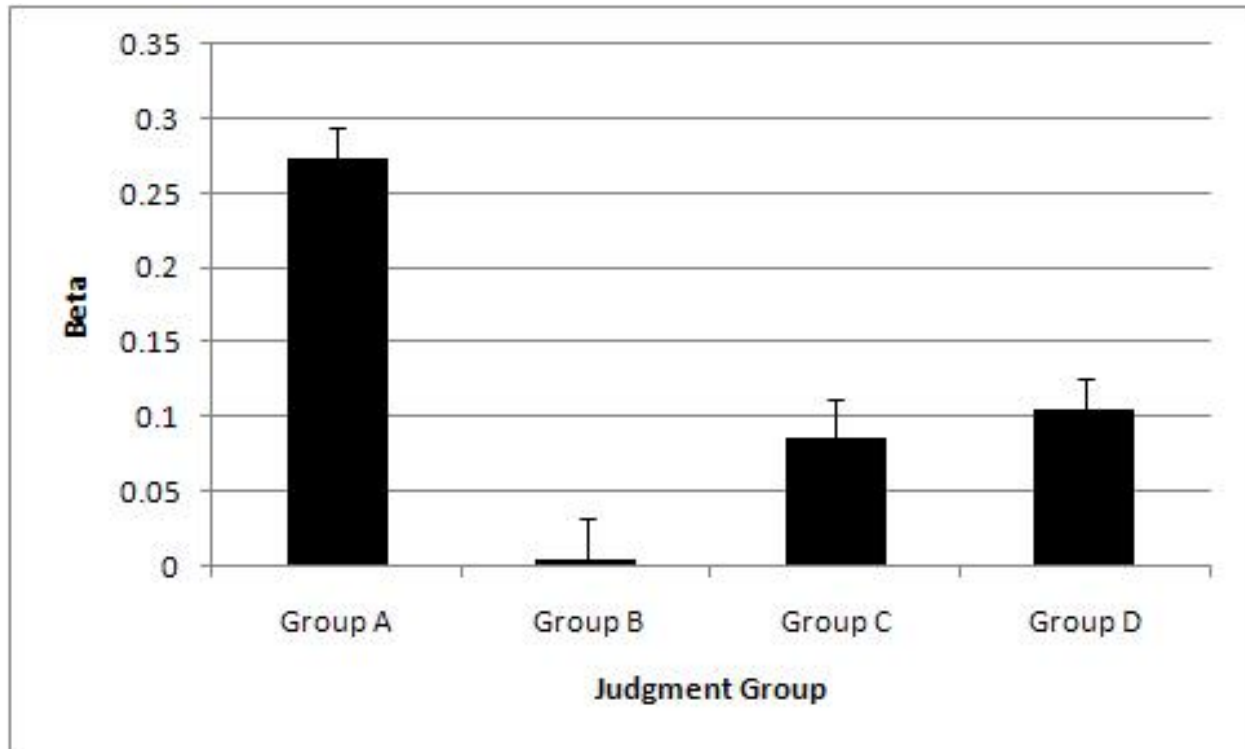associates or only a few (Group D, $F(2,22) = 1.559$, $p = .223$, $R^2 = .071$).

Figure 1. Betas for each judgment group with standard error bars. Group A did the traditional

judgment of memory task by rating word pairs on their associative relatedness. Group B tried to

free associate all the words related to a given cue word. Group C had to guess how many words

were related to a given cue word. Group D simply marked if they thought that the given cue

word had many associates or only a few.

Slopes were then averaged over participants for each group to compare the different

judgment of memory tasks. Figure 1 shows the average beta by type of judgment. A univariate

between-subjects ANOVA was analyzed to examine the differences in betas between the

judgment groups. As shown in Figure 1, there was a significant difference between the groups in

participants ability to judge associative relationships, $F(1, 129) = 23.989$, $p<.001$, $\eta^2 = .358$. A Tukey test was used for post hoc analyses. The traditional judgments of memory task showed a normal range beta from previous research ($M = .272$, $SE = .021$; Maki 2007a, 2007b), which was significantly higher than the other groups. Group D, who rated words as having many associates versus few associates had the next highest slope ($M = .104$, $SE = .022$), which was significantly different than Group B ($M = .004$, $SE = .028$), who tried to free associate words related to the cue word. Group C, who guessed at how many words were related to a cue word ($M = .086$, $SE = .027$), was not significantly different from Group B or Group D.

Discussion

Working memory is the ability to work with information and ignore other information (Kane et al., 2007). However, neither working memory, nor its close relative, fluid intelligence, was able to predict the ability to make associative judgments. Kahneman and Tversky (1973) have shown that people are not able to search through memory for examples. The judgments of associative memory task required participants to retrieve related word links in memory. Although participants varied in their ability to make judgments in these tasks, in several of the groups the slopes were very flat and close to zero. In the free association group, the slope was .004, which means that participants were able to name one or two associates for each cue word. One or two items in short term memory is well within the working memory span of most individuals. The ability to withdraw information from memory appears to be separate from the ability to work with the information that is already pulled up from memory.

The fan effect can explain why related information is difficult to retrieve from memory (Anderson & Reder, 1999). Participants were given sentences about different concepts, such as

lawyers, and required to remember which facts were connected to each concept. Anderson and Reder found that concepts with many relations were remembered less often than concepts related to only a few pieces of information. They argue that as a concept is activated, the activation spreads out in a fan like fashion to other related information. The activation is then split over the links in memory, which can make activation very small for a concept with numerous associations in memory. Because activation is small, the individual related concepts are difficult to recall.

The free association database contains the different strengths of relationships between words (Nelson et al., 2004). Many concepts have one very strong relationship, such as LOST-FOUND, which has a probability forwards strength of .74. Even if activation is spread across all related links, the strongest connection can still be pulled from memory. The rest of the associates have smaller probabilities with little activation, which makes them difficult to bring to short term memory. Other concepts, like COMPUTER, have many very small associative links, making the activation of any individual link negligible. As seen in the judgments of learning literature, the strength of relationship between associative word pairs predicts both recall and confidence judgments of learning (Koriat & Bjork, 2006).

The strengths of word pair associations and the fan effect can also explain why the slopes of judgments of memory are almost zero. All groups received instructions about the different number of associative relationships in memory, specifically that some words had many relationships while others had only a few. Therefore, participants knew, especially in the groups asked to guess at the number of relationships, that they were to separate words based on their number of perceived relationships. If the activation of associates was very weak, participants would not be able to retrieve enough examples from memory to accurately make a judgment about the number of related connections.

In Maki's (2007a, 2007b) previous work, judgment betas for associative database scores averaged between $\beta = .2$ and $\beta = .3$. The current study found similar results for the replicated judgment of associative memory task. This task was performed significantly better than the other three versions of the associative memory task. This task is different from the other three tasks in that participants are judging the strength of the relationship between two words instead of the number of connections for a single word or trying to retrieve associates for a single word. Apparently, when participants focus on one relationship rather than trying to determine many relationships, they are better at judging the characteristics of concepts. Participants are also significantly better at judging if words are connected to many concepts or only a few. This effect could be due to the nature of the high/low decision task. The decision is limited to two extremes instead of having to count the number of concepts that could be thought of quickly. Participants could simply search through initial activation to determine the cue word's properties.

Even in the best judgment of associative memory task, the participant slopes were nowhere close to one, which would indicate perfect discrimination capability. The current research tried to make this task easier by changing the judgment task to a free association or simple number estimation task. None of the changes made were able to improve judgments and actually decreased participant ability to make judgments of associative memory. Working memory was examined as an influence to differences in slopes across participants, but was not a significant predictor of any of the various judgment tasks.

References

Anderson, J., & Reder, L. (1999). The fan effect: new results and new theories. *Journal of Experimental Psychology: General, 128*, 186-197.

Arbuckle, T., & Cuddy, L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology, 81*, 126-131.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior, 19*, 450-466.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128*, 309-331.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237-251.

Kane, M., Conway, A., Hambrick, D., & Engle, R. (2007). Variation in working memory capacity as variation in executive attention and control. *Variation in working memory* (pp. 21-46). New York, NY US: Oxford University Press.

Koriat, A. (2008). When confidence in a choice is independent of which choice is made. *Psychonomic Bulletin & Review, 15,* 997-1001.

Koriat, A., & Bjork, R. (2005). Illusions of Competence in Monitoring One's Knowledge During

Study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 187-194.

Koriat, A., & Bjork, R. (2006). Mending metacognitive illusions: A comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 1133-1145.

Maki, W. (2007a). Judgments of associative memory. *Cognitive Psychology, 54*, 319-353.

Maki, W. (2007b). Separating bias and sensitivity in judgments of associative memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 231-237.

Maki, W., Krimsky, M., & Muñoz, S. (2006). An efficient method for estimating semantic similarity based on feature overlap: Reliability and validity of semantic feature ratings. *Behavior Research Methods, 38*, 153-157.

Maki, W., McKinley, L., & Thompson, A. (2004). Semantic distance norms computed from an electronic dictionary (WordNet). *Behavior Research Methods, Instruments & Computers, 36*, 421-431.

McRae, K., Cree, G., Seidenberg, M., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, 37*, 547-559.

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

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

& Computers, 36, 402-407.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and*

Vocabulary Scales. Section 2: The Coloured Progressive Matrices. San Antonio, TX:

Harcourt Assessment.

Turner, M., & Engle, R. (1989, April). Is working memory capacity task dependent?. *Journal of*

Memory and Language, *28*, 127-154.

Unsworth, N., Heitz, R., Schrock, J., & Engle, R. (2005). An automated version of the

operation span task. *Behavior Research Methods*, *37*, 498-505.