In a JAM: How Judgments of Associative Memory are Aided by Expertise

Erin M. Buchanan

Missouri State University

Erin M. Buchanan

Missouri State University

901 S National Ave.

Springfield, MO 65897

Phone: 417-836-5592

Email: ErinBuchanan@missouristate.edu

Abstract

Tversky and Kahneman’s (1973) is famous inquiry, “Are there more words with the letter K in the first position or the third?” highlights our difficulties with making judgments from memory. People find it very easy to think of words that start with K (kite, knife), but have a harder time with K in the third position (make, bake), which is the correct answer. While this frequency decision sounds innocuous, research in judgments of learning (JOLs) shows we have a serious need to understand how underlying sorting and judgment from memory is accomplished. Koriat and Bjork (2005, 2006) have shown that participant’s ratings of test performance are nearly unrelated to actual test performance. Here, participants created their own set of free association norms over several weeks time and were then asked to judge the associative strength of their word pairs. These judgments were then compared to real associative strengths (Nelson, McEvoy, & Schreiber, 2004), a control group, and a matched group. Participants who judged their own pairs were better than all other groups implying that judgments based on individualized generated memories are better than judging the collective norm, but are albeit still far from perfect.

In a JAM: How Judgments of Associative Memory are Aided by Expertise

People often are asked to judge the frequency or probability of events. If you are waiting on a friend who is late to lunch, you have to decide if (s)he is always late (high frequency) or never late (low frequency), which will determine how long before you start to worry about them. Such judgments are grossly inflated in many situations, and student learning is one of the pertinent applications of judgment research. For example, people estimate their competence at levels higher than those supported by their actual levels of skill (Dunning, Johnson, Ehrlinger, & Kruger, 2003). Often, judgments of learning (JOLs) are also inflated leading to inappropriate study strategies (Koriat & Bjork, 2005). Individuals who self-monitor study habits are found to be highly confident and poor at calibration in learning (Cutler & Wolfe, 1989). Thus, development of methods for remediation of inflated predictions is important for both theoretical and practical reasons (Koriat, 2008; Koriat & Bjork, 2006).

Word associations are a convenient domain in which to investigate inflated predictions and their solutions (Koriat, Fiedler, & Bjork, 2006; Maki, 2007a). Word associations are usually measured by the method of free association. In this task, participants are asked to give the first word that comes to mind (target) when presented with a stimulus word (cue). Averaged over many participants, the forward strength (FSG) of the association between words A and B is indexed by the conditional probability that word B is given in response to word A. Large databases are available that list the associative strengths of many thousands of cue-response pairs (e.g., Nelson et al., 2004). For example, the probability of *computer-program* appearing together is about 12% or 12 people out of a 100.

Studies of the inflation of predictions within the domain of word associations have shown that people typically over-estimate word relations, especially for weakly associated pairs. This tendency toward high ratings is most difficult to change (Koriat, 2008; Koriat et al., 2006; Maki, 2007a, 2007b; see also Nelson, Dyrdal, & Goodmon, 2005). For example, the inflation persists even when people are shown a list of common associates for a given cue, when people are given a cue and asked to think about alternatives (Maki, 2007a), and when people overtly generate their own list of associates (Koriat, 2008; Koriat et al., 2006); in all three cases judgments remained poorly assessed.

Because of the robustness of the inflation effect, finding a substantial reduction is important. Three different training methods have proven effective in reducing the inflation effect, such as mnemonic and theory based debiasing. However, these training methods have only influenced the overall overestimation bias factor, but not our sensitivity to the difference between low and high frequency events. Bias is the general tendency to inflate judgments upwards (we think everything is strongly related), while sensitivity is our ability to accurately tune our judgments. The current experiment tested how using individually normed frequencies affected our ability to make judgments from memory. Previous research has shown that participants cannot estimate what *others* would say when given a particular cue word (Maki, 2007a, 2007b; Buchanan, 2009; Foster & Buchanan, 2012; Valentine & Buchanan, *under review*). In a traditional judgment of associative memory (JAM) task, participants are asked to estimate how many college students out of a 100 would list the second *target* word when presented with a first *cue* word. Here, we had participants create their own frequency norms, and then they were asked how often they put words together. If these frequency memories are built over time (akin to distributed practice study skills), we may see a boost in judgment capacity over comparison groups. The following hypotheses will be tested:

* Hypothesis 1: Word frequency will match previous database frequency on an individual and overall participant level. This hypothesis will be examined with correlations.
* Hypothesis 2: (A) Free association database norms will be predictive of all group’s judgments. Simple linear regressions will be used to calculate the slope of judgments when compared to free association norms. These values will be tested against non-predictive slopes (i.e. *B* = 0) with single sample t-tests. (B) Furthermore, the individualized experimental group norms will be positively related to their judgments of associative memory, using the same procedure to test their slope values above zero.
* Hypothesis 3: Judgment ability will vary across groups, so that the experimental group should show better judgment ability when compared to their own norms over control, matched and experimental groups compared to free association norms. This hypothesis will be examined on slope values with a mixed factorial ANOVA and dependent t-test.

These hypotheses will highlight how expertise (interaction with word pairings and their frequency) affects judgment processing on the memory system. In a distributed practice study, one would expect judgments of learning to be high because participants know they have practiced items repeatedly, which would increase their frequency and likelihood of remembrance. Here, judgments of memory should also be high when items have been repeated over time (thus strengthening their connection in memory).

**Method**

**Participants**

Participants were recruited through the department of psychology’s undergraduate subject pool. Students were required to participate in research for their general psychology course, and some upper level courses allowed research participation for extra credit. The research project was displayed on the SONA system, which is online participant-credit management software. Participants choose research projects to complete based on availability and interest in the study’s posted abstract. The entire experiment was completed online, and each section took approximately 5 to 15 minutes. In the experimental group, fifty-five participants started the experiment, where *N* = 41 completed all experimental sessions and *N* = 14 did not finish all sections of the experiment. For the non-finishing group, the average number of sessions was *M* = 2.14 (*SD* = 1.17) with a range of 1 to 4 rating sessions. The comparison groups included fifty-two participants for the control group and forty-one participants for the matched group.

**Materials**

Stimuli were selected from the free association word norms by Nelson et al. (2004). The database contains different word-pair combinations, along with the probabilities of those pairings. For example, with the pair *steak-sirloin*, *steak* is the cue word that is paired with the target word, *sirloin*. Each cue word (the first word) has several different target words (*steak-cow, steak-sauce*). Cue words were selected with varying number of target combinations, but specifically words with more cue-target combinations (i.e. 10 or more). These words were selected so that there were many associated target words for participants to produce. Obviously, individual experience will influence free association, but cue words with many associates allow for many varied answers more so than cue words with one or two highly associated words. Twenty cue words were selected to use in the experiment. Target word selection is described below.

**Procedure**

***Experimental Group.*** This group of participants was given the chance to compare their own, actual pairing probabilities instead of others’ likely judgments. In the norming stage of the experiment, participants were given instructions on a free association task. The instructions described free association as the “first word that pops into your mind when you hear a *cue* word”. For example, many people talk about owning a *cat* and a *dog*, but we may also mention that it’s raining *cats* and *dogs* outside. These examples help portray free association as language *use* and not just language *meaning*, so that participants do not simply write only *fur, tails,* and *whiskers* when asked about *cats*. After these instructions, participants were given 20 cue words with four blanks each. For each cue word, they wrote the first four target words that come to mind, which was intended to create some variation in target words during the first stage. After they completed all the cue words, their answers were stored.

After a minimum of two days, participants were allowed to take the survey again. Participants were sent email reminders when they were allowed to complete the next section of the study. Each participant took the survey five times, and cue words were randomized with each presentation. These trials were averaged and probabilities were created for each cue-target pairing, similar to the Nelson et al. database. For example, over the course of five trials, participants may have listed many words for the cue *computer*, such as *mouse*, *screen*, *game*, *program*, *keyboard*, *data*, *fast*, etc. Each of those pairings will have a probability of occurring for that participant (i.e. they may have listed *screen* on all 5 surveys / 5 possible = 100%). From these pairings, 50 cue-target combinations were selected: ten word-target pairs from each possible probability (20%, 40%, 60%, 80%, and 100%). Participants were asked to estimate the probability of their combination of each cue-target pair. For instance, a participant might see the following: “When asked about *computer*, you listed the word *program*. What percent of the time did you put *computer* and *program* together?” After they finished the final survey, a debriefing of the project was shown.

***Control Group.*** Results froma separate control group were compared against the experimental participant judgment scores. Since each experimental participant’s final word pairs were different, a group of cue-target pairings was selected from the Nelson et al. database to match the experimental groups. The same 20 cues were used, and target words were mixed so that there was an equal number of low, medium, and high pairings. The control group was given the same instructions about a free association task, along with examples. Next, the rating task was explained as follows: “How many people out of a 100 would give the *target* (second) word when asked the *cue* (first) word?” Participants estimated the probability of word pair occurrence for the next 50 words after some practice trials. Several cues were repeated to create 50 word-pairs, to match the repetition in the experimental group.

***Control Matched Group.*** Lastly, a separate group of participants were used as comparison to the experimental group. These participants were randomly paired with an experimental participant. They were given the control group instructions about free association and the same instructions for the rating task. However, instead of judging randomly selected word-pairs, they were shown experimental participant normed word-pairs. This group will be used to examine the nature of stimuli on judgments, showing that participant judgments improved because of their interaction with words in the experimental group.

**Results**

**Descriptive statistics (Hypothesis 1).** The data were tabulated in several different ways, as shown in Table 1. First, the average number of target words per person and mentions was calculated. This value is the average number of cue-target combinations that were mentioned once, twice, etc. calculated within subjects. In total, across all five testing times, participants mentioned approximately 206 different target words, which were mostly only mentioned once. The average number of targets listed by stimuli and mentions show that the average number of words listed for each of cues was 423, again with the majority of targets mentioned once. Interestingly, a small bump in average number of words was found for five mentions. A large majority of once mentioned targets portrays that participants are not simply listing the same four words on each cue and trial, but instead are listing large number of unique words for each cue word. However, there are clearly targets that have strong associations or you might expect the average number of words per mention to continue to decrease across mentions.

Second, 8,457 total cue-target pairs were created across all participants, cues, and experimental sessions. The percent of targets mentioned once was the largest with a small increase for the number of targets mentioned five times. 7,635 of these target words were unique within participants. Target words were allowed to overlap across participants but not within participants for calculation; for example, a participant might have written *achieve* to both *challenge* and *difficult*. The average number of unique targets per participant was *M* = 186.22 (*SD* = 37.66). 3,252 unique targets were found by stimuli applying the same criteria as participants (no-repeats within stimuli, but allowed across stimuli). The average number of unique targets by stimuli was *M* = 162.60 (*SD* = 29.49). When all targets were considered regardless of stimuli or participant, 2,277 unique targets were listed.

These pairs were then compared to the Nelson et al. (2004) database where 29.40% percent of participant pairs had non-zero normed values. When unique cue-target pairs are considered (i.e. no overlaps between participants), approximately 15.50% of unique pairs had non-zero values in the Nelson norms. The correlation between participant normed values (20%, 40%, 60%, 80%, 100%) and forward strength was *r* = .44, *p*<.001. Further, if we calculate the number of cue-target pairs that exist for the stimuli set, 75.96% of the non-zero possible pairs were listed across participants implying that most targets were listed at least once. The correlation between the percent of participants who list a pair at least once and forward strength was *r* = .68, *p*<.001. This percent is the probability of each cue-target pair, as if they had been collected in the normal free association style, with one target per cue (collective forward strength). Lastly, the correlation between the percent of all mentions of the target (i.e. number of mentions / number of times the target could have been mentioned) was *r* = .79, *p*<.001. These statistics suggest that while many unique words were obtained outside of the Nelson norms, there was a great deal of expected overlap within the current data and previously normed words; thus supporting Hypothesis 1.

**Dependent variable calculation.** In this experiment, the dependent variable was considered the slope value calculated when using a simple linear regression to predict participant judgments. First, each participant’s judgment scores were matched to the Nelson et al. (2004) free association norm data. As described above, some pairings created by the experimental groups were not present in the Nelson et al. norms, so these data were discarded when calculating the slope values for that group. Judgment and free association data was scaled to a 100 point system, so that all B values are comparable across groups and to previous research (Maki, 2007a, 2007b). Finally, the free association norms were used as the independent variable to predict participant judgment scores, creating a measure of sensitivity to the difference between lowly and highly related word pairs. This procedure created three norm matched variables call FA norms below in tables: (1) control group ratings, (2) experimental group ratings, and (3) matched group ratings. Second, experimental group participant norms were matched to both the experimental group and matched group ratings on judgment pairs. Their norms were used to predict ratings in a simple linear regression, again scaled to 100 points for comparison. This procedure created two sets of participant normed sensitivity slopes: (1) experimental and (2) matched.

For all analyses described below, control groups were considered a between subject’s variable for comparison to matched and experimental groups because both the participants and stimuli were different. The experimental and matched groups were considered repeated measures variables because the stimuli and norms used to create the dependent variable were exactly the same (i.e. the matched group was yoked). Therefore, a main difference between experimental and matched groups was the experience with creating the word set, which was the primary goal of this analysis. Further detail about the motivation of this analysis procedure is listed with the individual hypotheses. All assumptions were analyzed for individual analyses, including outliers, and no problems were found.

**Hypothesis 2.** Hypothesis 2 examined if participants were sensitive to the strength of relation between word pairs and could discern the difference between lowly and highly related word pairs. Table 2 shows the average slope values for each group-normed association combination. To test Hypothesis 2, single-sample t-tests were used to compare group slope values against zero. If slope values were significantly greater than zero, this finding would indicate that groups were able to complete the judgment task and predict the relationship between word pairs better than chance. Previous studies have shown slope values ranging from 0.2 to 0.5 (Maki, 2007a, 2007b, Nelson et al., 2005) when participant scores are compared to the Nelson et al. (2004) free association ratings. As indicated in Table 2, all single sample values were significantly greater than zero at *p* < .001. An examination of the mean and Cohen’s *d* values show that the slopes for control, matched, and experimental-free association normed groups were large effect sizes matching previous literature (i.e. slopes averaging around 0.20). As predicted in Hypothesis 2B, the experimental-participant norm group showed a larger mean, with a very large effect size when compared to a non-sensitive slope of zero. Therefore, Hypothesis 2 was supported, showing that all groups were able to judge word pairs better than random guessing. Interestingly, matched groups compared to participant norms were also significantly greater than zero at *p* <.001. This finding seems to indicate that matched participants were able to judge the strength of pairs created by others, at about the same level as when compared to free association norms and control groups (however, the difference was tested below).

**Hypothesis 3**. Next, this hypothesis examined the relationship between group slope values to show that even though norms are predictive of participant judgments, the experimental group ratings of their own norms are higher than matched or control group norms. First, control groups were compared to all other groups using independent t-tests, followed by experimental and matched groups analyses using repeated measures ANOVA. Control groups were tested with similar but different word pairs on their associative judgments, while experimental and matched pair groups were tested on the same word pairs. Therefore, even though participants were different in the experimental and matched conditions, their data was tied to the same stimuli and considered repeated measures data. Furthermore, the comparison between free association and participant normed data was repeated on the same data, making a 2 (condition) X 2 (norm comparison) repeated measures ANOVA analyses appropriate. These analyses will indicate if the experimental group’s experience with the word set allowed them to more accurately judge word pairs over a matched person’s judgment of the same word pairs. This interaction with the word set should also improve their judgments over free association norms predicting their judgment scores.

Control group slopes were not expected to differ from matched groups or experimental groups when compared to free association norms. This finding was expected because in these groups, participants are judging the strength of word pairs in the same manner as the control group (matched groups procedure) or being compared to the population average free association (experimental groups), which mimics the control group task. Table 3 portrays the t-test differences between groups, in which these findings are shown. Control groups were actually significantly greater than matched groups when compared to the participant norms, indicating that when participants are asked to rate how many people out of a 100 would list the second word to the first, participants are better when compared to the normed average than rating individual made averages. Lastly, control group slopes are much lower than experimental groups when compared to participant norms, as expected since experimental groups are judging their own pairs instead of an averaged probability with the free association norms.

A 2X2 repeated measures ANOVA showed groups differences overall, *F*(1, 40) = 25.65, *p* < .001, ɳ2 = .39, as well as the comparison between participant norms and free association norms, *F*(1, 40) = 24.58, *p* < .001, ɳ2 = .38. However, the interaction between norm comparison and group was significant, *F*(1, 40) = 41.09, *p* < .001, ɳ2 = .51, so only this effect will be analyzed with dependent t-tests for post hoc. These t-test values are shown in Table 3. Experimental group slopes were significantly larger than matched group slopes when they are compared against the experimental group norms, which show that the creation of individualized norms allows participants to judge relationships better than judging other participant word pairs. Further, these experimental group participant norms are higher than the free association counterpart for both matched groups and experimental groups. Therefore, we can attribute the improved experimental group performance to previous interaction with word-pairs instead of just “easy” stimuli, otherwise these groups would all show the same slope values (i.e. groups would all show a small sensitivity to strength as seen in the control group analysis). Lastly, matched group comparisons were not significantly different, further supporting the experimental group’s interaction hypothesis. This finding is important because a stimuli specific finding would show that matched group participants norm slopes should be greater than the free association norm slopes (i.e. the words created by the experimental group are just somehow obvious at the differences in low and high strength pairs).

**Discussion**

In viewing these findings, it appears that participants can judge the associative strength between word-pairs, and quite well when given their own associative norms to judge. The experimental group was better at judging low and high frequency relationships when compared to a matched group of participants and a control group of participants. The interaction with the word-pairs over repeated trials improved performance for judgments, indicating that expertise has a positive effect on judgment performance. This result is not too unexpected as much research shows that experts are better at working memory tasks within their expertise (Chase & Simon, 1973; Ericsson & Delaney, 1998), as well as better suited to work within long term memory (Ericsson & Delaney, 1999). However, previous research on JAM showed that participants were not able to improve their performance when given practice and feedback with judgments (Maki, 2007b; Koriat & Bjork, 2006), which is also a large component of expertise.

Potentially, the difficulty in judgments could be embedded within the experimental design. Participants rate what a collection of 100 college students have said for a cue word, which is how the original free association norms were collected. In theory, forward strength probabilities are the collective likelihood of obtaining a target word when given a cue word (Nelson, McEvoy, & Dennis, 2000). Why should we expect participants to be able to guess at other’s free association when this study showed that many unique words were created in the norming process? Two reasons: (1) simply put, people estimate probabilities of events *all the time*, and (2) the current study showed that much of these collective norms were reproduced across the dataset. The popular game show Family Feud had contestants guess the most likely answers to different cued categories and would have been very boring indeed if people could not estimate a common answer to a cue. During normal day we estimate all sorts of events: how long will it take to get to work, did I turn off the stove (how likely was I to remember to turn it off?), how long should I wait for a friend for lunch, have I studied enough for a test, etc. All of these functions require an estimate based on previous knowledge of probabilities of events.

The daily practice of estimation does not seem to make us any better at the actual estimation. Koriat (2008) and Bjork’s (2005, 2006) research on metacognitive control implies that participants can provide better estimations of how well they have learned something (and subsequent performance on an exam) but only with increased awareness of the pitfalls of foresight bias. This work has shown that participants are better at judging their own normed probabilities, but even these judgments are not perfect. A very sensitive judgment slope would be close to a *B* of one. Here, the average slope for the experimental group when compared to their own norms was *B*  = .54, and ranged from *B* = -.18 to *B* = .80 (an average similar to Nelson et al.’s (2005) studies on judgments). Therefore, while it is clear that judging one’s own memories improves estimation, judgments are still less sensitive to the actual frequency of events. We can easily tell students to “just keep studying” to alleviate the problem of overestimation of learning and performance, other judgment estimations may not be so easy to fix. Expanded studies may wish to investigate the role of the stimuli (e.g. word frequency) on judgments, which may show that our interaction with words and their relations has a mitigating effect on our estimations of frequency.

**References**

Buchanan, E. (2009). Access into memory: differences in judgments and

priming for semantic and associative memory. *Journal of Scientific Psychology*,1-8.

Chase, W.G., & Simon, H.A. (1973). Perception in chess. *Cognitive Psychology, 4,* 55-81.

Cutler, B.L., & Wolfe, R.N. (1989). Self-monitoring and the association between confidence and

accuracy. *Journal of Research in Personality, 4,* 410-420.

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their

own incompetence. *Current Directions in Psychological Science*, *12*, 83-87.

Ericsson, K. A., & Delaney, P. F. (1998). Working memory and expert performance. *Working*

*Memory And Thinking: Current Issues In Thinking And Reasoning*, 91.

Ericsson, K. A., & Delaney, P. F. (1999). Long-term working memory as an alternative to

capacity models of working memory in everyday skilled performance. *Models of working*

*memory: Mechanisms of active maintenance and executive control*, 257-297.

Foster, L., & Buchanan, E. (2012, *in press*). Is insensitivity limited to one word-pairing or four?, *Journal of Psychological Inquiry.*

Koriat, A. (2008). Alleviating the inflation of conditional predictions. *Organizational Behavior*

*and Human Decision Processes*, *106*, 61-76.

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

during study. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *31*,

187-194.

Koriat, A., & Bjork, R. A. (2006). Mending metacognitive illusions: A comparison of

mnemonic-based and theory-based procedures. *Journal of Experimental Psychology:*

*Learning, Memory, & Cognition*, *32*, 1133-1145.

Koriat, A., Fiedler, K., & Bjork, R. A. (2006). The inflation of conditional predictions. *Journal*

*of Experimental Psychology: General*, *135*, 429-447.

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

Maki, W. S. (2007b). Separating bias and sensitivity in judgments of associative memory.

*Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 231-237.

Nelson, D. L., Dyrdal, G. M., & Goodmon, L. B. (2005). What is preexisting strength?

Predicting free association, similarity ratings, and cued recall probabilities. *Psychonomic*

*Bulletin & Review*, *12*, 711-719.

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*, 402-407.

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

measure?. *Memory & Cognition*, *28*(6), 887-899.

Tversky, A., & Kahneman, D. (1973). Availability: a heuristic for judging frequency and

probability. *Cognitive Psychology*, 5, 207–232. doi:http://dx.doi.org/10.1016%2F0010-

0285%2873%2990033-9

Valentine, K.D., & Buchanan, E.M. (*under review*). JAM-boree: An Application of Observation

Oriented Modeling to Judgments of Associative Memory.

Table 1

*Descriptive Statistics of Experimental Group Norms*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Number of Mentions | | | | |  |
|  | 1 | 2 | 3 | 4 | 5 | Total |
| Participant *M (SD)* Targets | 119.76 (39.25) | 31.10 (10.09) | 18.27 (5.87) | 16.41 (4.55) | 21.25 (8.31) | 206.27 (40.23) |
| Stimuli *M (SD)* Targets | 245.50 (46.37) | 63.75 (11.01) | 37.45 (6.13) | 33.65 (5.43) | 42.50 (15.15) | 422.85 (43.40) |
| Percent Total Targets | 58.06 | 15.09 | 8.84 | 7.96 | 10.05 | 8457 |
| Percent Unique Targets by Participant | 59.02 | 14.97 | 8.74 | 7.79 | 9.48 | 7635 |
| Percent Unique Targets by Stimuli | 74.45 | 11.93 | 6.37 | 3.84 | 3.41 | 3252 |

*Note*. Numbers in parentheses are standard deviations.

Table 2

*Slope and Intercept Values for Normed and Participant Created Association Values Predicting Participant Judgments.*

|  |  |  |  |
| --- | --- | --- | --- |
| Group | *B (SD)* | *t* | *d* |
| Control | 0.28 (0.17) | (51) 11.85 | 1.65 |
| Experimental – Participant Norms | 0.54 (0.18) | (40) 19.43 | 3.00 |
| Experimental – FA Norms | 0.26 (0.20) | (40) 8.46 | 1.30 |
| Matched – Participant Norms | 0.20 (0.17) | (40) 7.62 | 1.18 |
| Matched – FA Norms | 0.23 (0.23) | (40) 6.42 | 1.00 |

*Note*. FA = free association for the Nelson et al. (2004) norms. For slope values, parentheses indicate standard deviations for the means. For t-tests, degrees of freedom are listed in parentheses. All t-test values significant at *p*<.001. *d*-values are Cohen’s d for single sample t-tests.

Table 3

*T-tests and Post Hoc Analyses for Hypothesis 2*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Mdifference* | *t* | *p* | *d* |
| Control Group Analysis |  |  |  |  |
| Matched - FA Norms | 0.05 | 1.19 | 0.24 | 0.25 |
| Matched - Participant Norms | 0.08 | 2.28 | 0.03 | 0.47 |
| Experimental - FA Norms | 0.02 | 0.53 | 0.60 | 0.11 |
| Experimental - Participant Norms | -0.26 | -7.14 | 0.001 | -1.49 |
| Interaction Analysis |  |  |  |  |
| Participant Norms - Exp to Match | 0.34 | 8.09 | 0.001 | 1.26 |
| FA Norms - Exp to Match | 0.31 | 6.70 | 0.001 | 1.07 |
| Exp Participant to FA Norms | 0.28 | 8.20 | 0.001 | 1.27 |
| Match Participant to FA Norms | -0.03 | -0.89 | 0.38 | -0.13 |
| *Note*. Control groups *df* = 91, experimental match groups *df* = 40. | |  |  |  |