

Syracuse University

SURFACE

Syracuse University Honors Program Capstone Projects Syracuse University Honors Program Capstone Projects

Spring 5-1-2006

Memory & Cognition: What difference does gender make?

Donna J. Bridge

Follow this and additional works at: https://surface.syr.edu/honors_capstone



Part of the [Cognition and Perception Commons](#), [Cognitive Psychology Commons](#), and the [Other Psychology Commons](#)

Recommended Citation

Bridge, Donna J., "Memory & Cognition: What difference does gender make?" (2006). *Syracuse University Honors Program Capstone Projects*. 655.

https://surface.syr.edu/honors_capstone/655

This Honors Capstone Project is brought to you for free and open access by the Syracuse University Honors Program Capstone Projects at SURFACE. It has been accepted for inclusion in Syracuse University Honors Program Capstone Projects by an authorized administrator of SURFACE. For more information, please contact surface@syr.edu.

Memory & Cognition: What difference does gender make?

Donna J. Bridge

Department of Psychology

Syracuse University

Abstract

Small but significant gender differences, typically favoring women, have previously been observed in experiments measuring human episodic memory performance. In three studies measuring episodic memory, we compared performance levels for men and women. Secondary analysis from a paired-associate learning task revealed a superior ability for women to learn single function pairs (i.e. words that are represented in only one pair), but performance levels for double function pairs (i.e. pairs that contain words that are also used in one other pair) were similar for men and women. We also reanalyzed data from a recognition experiment that used pictures as stimuli, and discovered an enhanced propensity for women to recollect the test probes in comparison to men, but familiarity based judgments had minimal differences between genders. A prospective study was conducted in order to compare the effect of gender on a delayed free recall task that included basic arithmetic problems as part of the distractor task. Implications for gender differences are discussed with regard to biological factors involving estrogen, and relevant social factors.

Contents

Introduction	1
Delayed Free Recall	2
Paired-Associate Learning	10
Recognition Memory	19
Discussion	25
References	32

Supported by MH069938-01 from the NIH to PI Marc W. Howard. The author acknowledges the continuous support and assistance from Marc Howard, Jennifer Probyn, Vijay Venkatadass, and Yaofei Zhang. Additional thanks to the remaining members of MEMlab for providing technical and moral support.

Introduction

Although gender differences have been observed in some episodic memory domains, the nature of these differences has not been well established or understood. The presence of gender differences in episodic memory performance has been noted under various laboratory conditions involving several different tasks, including immediate and delayed free recall of words (Herlitz, Nilsson, & Backman, 1997) and recognition of words (Herlitz et al., 1997) and faces (Herlitz & Yonker, 2002; Lewin & Herlitz, 2002), all favoring women. The observed differences have been reliable, but the effect sizes are often small and typically require a large subject pool to see an effect, (Herlitz et al., 1997; Herlitz & Yonker, 2002; Lewin & Herlitz, 2002; Kramer, Yaffe, Lengenfelder, & Delis, 2003). Nonetheless, the variable performance accounted for by gender classification has been consistently observed in tasks measuring episodic memory performance. Therefore, we find it intrinsically interesting to the study of memory to examine the role that gender plays in memory processes and its subsequent effects on performance.

By using advanced techniques to describe episodic memory processes in great detail, we have begun to unveil subtle influences of gender. Within three episodic memory domains, data from four different experiments have been assessed on the basis of gender. We compared memory performances of men and women with two verbal learning tasks, by implementing a delayed free recall experiment and reanalyzing data from a paired-associate study, see (Howard, Jing, & Probyn, Submitted). Further insight was provided by secondary analysis of two non-verbal studies measuring recognition memory, that relied on the use of pictures to serve as stimuli, (Howard, Bessette-Symons, Zhang, & Hoyer, 2006; Schwartz, Howard, Jing, & Kahana, 2005).

With the present study, we attempt to describe the influence of gender that has been observed in various episodic memory procedures, by utilizing sophisticated methods to describe underlying processes. As a comparative feature, we examine

the differences elicited by a basic math assessment, and provide speculation about the potential source of these cognitive gender differences.

Delayed Free Recall

We are interested in mapping out the unique organization of previously learned items or events in memory. Free recall designs provide a useful way to assess the processes underlying episodic memory retrieval. In a free recall experiment, subjects are presented with a list of words individually during a study phase. Following a variable delay after the study period, subjects are instructed to recall as many words from the list they can remember in any order. Because subjects are not required to recall the items from the list in the same sequence they were originally presented, the output positions of the remembered words are unique to individual subjects. Even though there is no restriction imposed on the recall period regarding order of output, certain patterns of remembering are consistently observed in free recall experiments. For instance, given a list of words in an immediate free recall task, the probabilities of recalling words located in the beginning and the end of the list are typically greater than for words from intermediate serial positions. The heightened probability of recall for words from early serial positions is referred to as the primacy effect. The most often recalled words in a free recall experiment tend to be items that occupy the final serial positions on the list. The recency effect describes this robust property observed in free recall paradigms. Because the items located toward the end of the list were the most recently presented, memory for these items may be stronger than words with preceding serial positions. Generally, both of these properties are pronounced in free recall designs; however, it is possible to reduce the effect of primacy by prohibiting rehearsal during the task, and the recency effect can be minimized by implementing a single delay period following presentation.

Methods. We recruited 67 women and 57 men from an introductory psychology class at Syracuse University, all of whom participated in the study for course credit.

We collected gender and ethnicity information from all participants, as part of the documentation we are required to issue by the NIH. A total of 119 subjects were used in our final analysis due to insufficient performance on the distractor task by five subjects, discussed in further detail below. In the course of an hour, subjects completed 18 trials of the task, each of which began with a study phase, immediately followed with a distractor task, and finally ended with a free recall period.

For each trial, participants were instructed to study a list of 25 words, derived from the Toronto word pool (Friendly, Franklin, Hoffman, & Rubin, 1982), and then later asked to recall any items from the list they could remember following a brief delay of 30s. Each word was visually displayed on the screen for 1100 ms, in which time subjects were required to identify the noun on the screen as a 'concrete' or 'abstract' word by pressing either the left or right control keys on the keyboard. This served as an orienting task to assure the participants' attention to the task. The inter-presentation interval (IPI), the lapse of time in between the presentation of each word, was set to 200ms. During the delay, subjects were prompted with a series of basic arithmetic problems involving the addition of three single positive integers. Answers were typed in using the number keys on the keyboard, and the number of math problems varied with each trial depending on the subject's response time for each equation. After the distractor period, participants were prompted with a beep that indicated to them that they should begin recalling any items from the list by speaking the words into the headset provided. Each recall period lasted 60s. Completion of the 18 trials was followed by a surprise final free recall period, during which time participants attempted to recall any words from any of the lists that had been presented to them during the experiment. Five minutes were allocated to the final free recall period.

Results. Several measurements are typically used to assess patterns of learning in free recall experiments. The first measurement we used to assess performance was a serial position curve (see Figure 1). A serial position curve clearly demonstrates

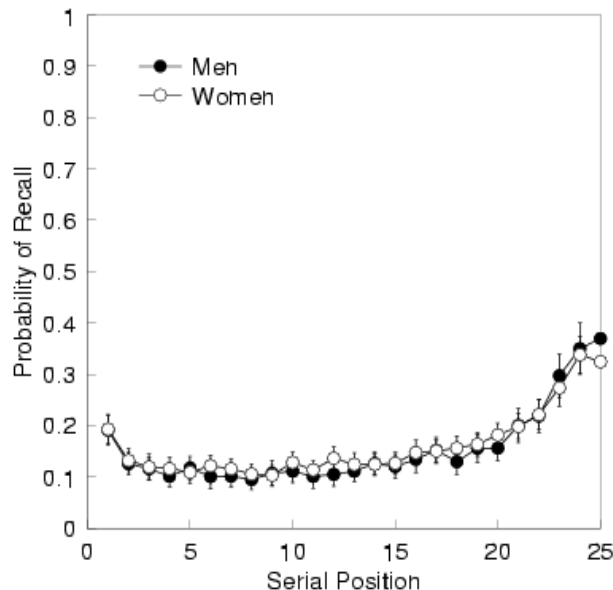


Figure 1. Serial position curve for men and women. Error bars reflect 95% confidence intervals.

the impact of recency and primacy on the recall period, by depicting the probability of recalling an item from the list relative to its serial position. Not only does the serial position curve highlight the recency and primacy effects, but it also serves as an overview to examine the probability of recalling items distributed across all serial positions. Figure 1 displays the cumulative probability of correctly recalling an item from a list collapsed across trials. Using the serial position curve, we are able to determine the average number of recalls per trial, by adding up each subjects' probability of recall for every serial position. We observed minimal differences between average number of recalls, with men recalling approximately 3.9 and women averaging 4.03 words per trial, $t(117) = -.62$ $p > .5$.

Although the serial position curve is a useful tool to analyze free recall data, it is limited in its ability to unveil subtle patterns of retrieval. A graph of the probability of first recall is useful for providing additional insight into the complexities that contribute to free recall paradigms. The probability of first recall (PFR) graph

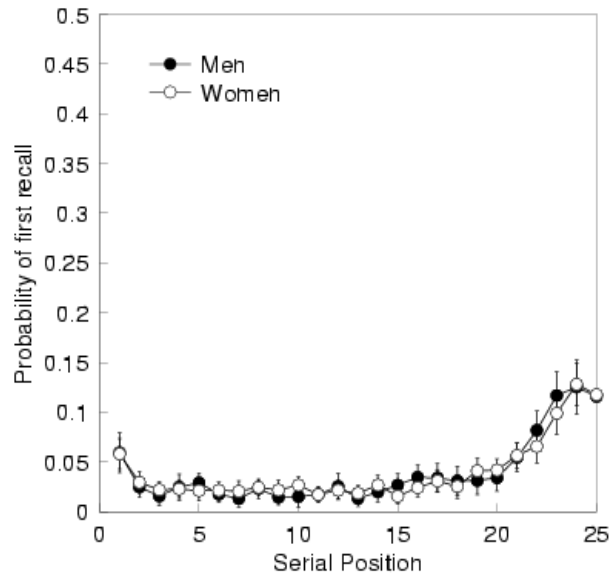


Figure 2. Probability of First Recall. Error bars indicate 95% confidence intervals.

is essentially a serial position curve, but it only considers the very first item recalled, and neglects subsequent output positions (Howard & Kahana, 1999, 2002). In immediate free recall, the PFR graph reinforces the pronounced recency effect depicted in standard serial position curves. Its main contribution, however, is its novel ability to unveil the heightened probability of recalling items from the end of the list in early output positions. The PFR curve (Figure 2) depicts similar results to the serial position curve, with men and women demonstrating indistinguishable probabilities of recall for the first output position.

It is possible to attenuate the pronounced recency effect observed in immediate free recall by increasing the lapse of time between presentation and recall, while simultaneously implementing a distractor task to disrupt rehearsal. The diminished recency effect can be observed in either the PFR or the serial position curve. Both curves highlight the primacy effect that remains in a free recall task with a delay, and it displays a reduced recency effect that is comparable in size to the primacy effect. We used a delayed free recall task in order to divert subjects' tendency to mainly

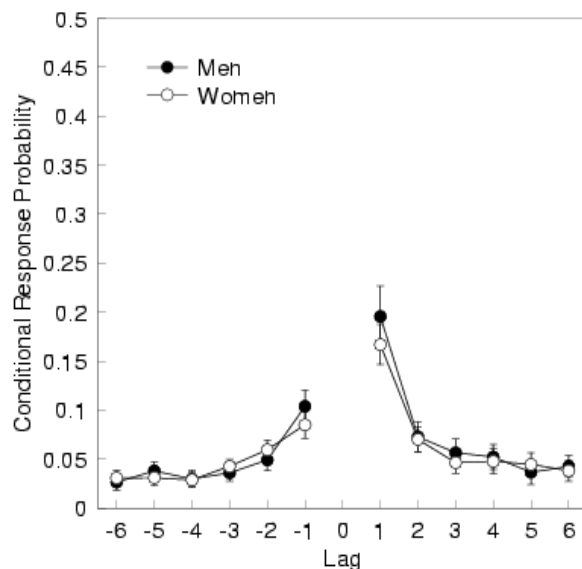


Figure 3. Conditional Response Probability curve for men and women. Error bars are 95% confidence intervals.

recall words from the end of the list, so that we would be able to focus primarily on the transitions that occur between items dispersed throughout the list during recall. An additional measurement is used to analyze specific transitions between words according to their output positions and relative to original serial positions at which they were presented. We constructed a conditional response probability (CRP) curve to describe the temporal relationship between words from the list with consecutive output positions. The CRP curve captures the unique organization of items in memory based on the two properties of contiguity and asymmetry. Figure 3 shows the probability of recalling an item from the list set as a function of lag, where lag is the relative difference between serial positions of subsequently recalled items. So, the relationship between items that are recalled with succeeding output positions is depicted by the distance between their serial positions on the list. A small lag between successively recalled items means that they were originally presented in similar contexts of occurrence based on time and location. Items recalled with ensuing serial positions are denoted by positive lags, while negative lags represent

preceding items on the list. For instance, suppose a subject is presented with a list of words, denoted here by letters, A - B - C - D - E - F. During the recall period, if the subject responds with D and then follows that recall with E, the lag would be indicated by +1. If the subject recalled C after D instead of E, the lag would be assigned a -1.

The weight of the associations presumably made during the encoding state are revealed during retrieval, and rely heavily on the amount of contextual overlap between the items. This is clearly demonstrated by the heightened conditional response probability of items with small absolute lags and a diminishing probability of recall as the lag increases. Interestingly, the CRP curve also exposes the peculiar property of asymmetry of associations. The CRP curve reaches its maximum value with a lag of +1, while a lag of negative -1 yields a drastically diminished probability of recall, but its value is still relatively high in comparison to probabilities with much greater absolute lags. The CRP curve (see Figure 3) provides the probability of recall relative output position, collapsed across output positions. We compared probabilities of recall with a lag of +1 and -1 but did not observe any distinguishable differences.

We examined the data from the delayed free recall experiment using a serial position curve, PFR curve, and CRP curve. These three tools permitted us to describe some the intricacies that underly memory retrieval. We compared men's and women's performances, but we did not observe any significant effects of gender with any of the measurements we utilized for this study.

Math Distractor. We initially examined the distractor analysis to confirm that subjects were following the directions of the task and not simply spending the 30s delay rehearsing words from the previous list. Our analysis contained several measurements describing each subject's performance, including the total number of math questions answered correctly and incorrectly, overall proportion correct, average reaction time, and variable reaction time. We decided to exclude those participants

who were not appropriately completing the task. Participants that exhibited variable reaction times greater than three standard deviations from the overall mean (> 4989 ms) were cut from the analysis, as well as those whose proportion correct was more than three standard deviations away from the mean ($< .77$ correct). A total of five subjects (all men) were excluded from our final analysis.

Interestingly, we discovered significant gender differences in the math distractor task. We used *t-tests* to compare reaction times, proportion correct, and total number of correctly answered equations. Overall, men and women answered the same proportion of math problems correctly (.93 vs .92 respectively), $t(117) = 1.18$ $p > .2$. Even though percentage correct was nearly equivalent, we discovered large advantages for men regarding total number correct and mean reaction time. Men were able to answer significantly more equations correctly $t(117) = 3.05$ $p < .003$, and at a faster rate $t(117) = -2.4$ $p < .02$, than women. Figures 4a and 4b display the distribution of total correct responses for men and women.

Distractor Discussion. Even though the cumulative number of questions answered correctly was higher for men, the level of accuracy was equivalent for both groups. Men may have been able to answer more math equations overall because their mean response times were significantly faster than women's. The ability to correctly answer basic math problems does not seem to differ across genders; however, since response time is a standard of measurement that is considered when assessing performance levels, men appear to be superior on this particular mathematical task.

Our examination of the detailed distractor analysis led to an unexpected influence of gender. The implementation of a distractor task in a delayed free recall experiment is primarily to divert subjects from rehearsal. The magnitude of the distraction produced by the task should be relatively equal among participants. Because of the divergent influence of gender on performance for the math task, we cannot assume that both groups were equally affected by the distractor task, and therefore men's and women's performances on the memory task may have been dif-

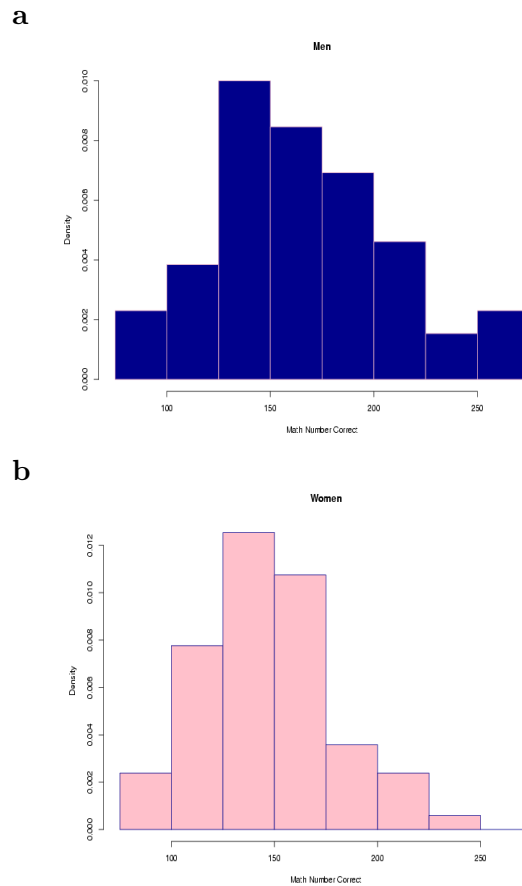


Figure 4. Distribution of total number of correct responses made by men and women. **a.** Distribution for men. **b.** Distribution for women.

ferentially influenced by the math problems. In future memory studies that do not take gender into consideration, the type of distractor task should be reevaluated so that all participants are comparably distracted.

The prevalence of stereotype threat may have an impact on some women's math performance. Stereotype threat refers to the psychological threat experienced by individuals who fear being categorized by negative stereotypes of a group they are affiliated with (Pronin, Steele, & Ross, 2004). Stereotype threat has the ability to affect performance of individuals in such a way that it actually reinforces the negative stereotype associated with the individual's group. Findings from some studies have suggested that mathematical evaluations can induce the negative influence of

stereotype threat on women's performances (Pronin et al., 2004; Johns, Schmader, & Martens, 2005). Presumably, the stress evoked by stereotype threat may provoke women to doubt their mathematical proficiency, causing them to exhibit slower response times than men. This stress may be a consequence of the broadly supported assumption that men are naturally more inclined to succeed in the realms of math and science.

The distractor analysis provides interesting insight into the standards by which we measure performance. For instance, standardized tests presumably measure mathematical ability and comprehension, but the strict time constraint imposed on these examinations may contribute to the divergent test scores exhibited by men and women. Because mathematics is an explicitly gendered domain of academia, where the achievements of men are greatly encouraged and a high aptitude is necessarily predicted at a very young age, women are subject to oppositional group stigmas (Pronin et al., 2004; Johns et al., 2005). Consequently, women who are affected by this stigmatization may demonstrate performances that reflect presupposed expectations about their individual abilities that stem from negative generalizations implicated by their particular gender group affiliation. The additional stress experienced by some women may induce doubt and anxiety that is exposed by their comparatively lengthy response times in relation to men.

Paired-Associate Learning

Understanding the underlying processes that determine memory performance is essential to break down the observable differences in memory between genders. Rather than measuring episodic memory performance by solely focusing on men's and women's aggregate correct recalls in verbal learning tasks, our analysis has been primarily devoted to the mechanisms underlying the retrieval of concrete information. In order to do this, we have assessed the learning of words by examining the associations formed between items by using a paired-associate learning task.

Associative memory can be described on multiple levels by assessing the learning of two pair types, double function and single function pairs. Single function pairs are composed of unique words that are used in only one pair, as seen in Figure 5a. Double function pairs consist of words that are used in two separate pairs (Primoff, 1938), as seen in Figure 5b. The words used in the double function pairs are part of a linked list; each word is matched with two unique items from the double function list and serves as a stimulus for one word and a response for another word during the test phase. For example, *C-D* and *B-C* are two pairs that share a similar item, *C*, that plays a specific role in each pair; as the correct response for the cue *B* and the probe for item *D*. Two particular pairs on the list do not follow the constraints of either pair type; the first and last pairs on the double function linked-list. The first and last pairs on the double function list *A-B* and *H-I* are distinguished from the other pairs, because *A* and *I* are represented in only one pair. Therefore, the learning of these two pairs is disregarded from our analysis of single function and double function pairs.

In this type of experiment, subjects are randomly presented with a list containing both pair types. Pairs are presented in an arbitrary order, however double function pairs containing a shared item are never presented consecutively. This constraint provides a means to assess the construction and strength of associations formed without the presence of direct temporal proximity between items in separate pairs. Because each double function item is a member of two separate pairs, recall is more difficult for this pair type than it is for single function pairs. This is presumably due to the presentation of each double function item in two distinct, yet overlapping contexts of occurrence (Howard et al., Submitted).

Intrusions. Our analysis takes into account all possible responses made to a cue item, including correct recall and intruding items. Although there is only one correct response to each stimulus, there are several different types of intrusions one could make in response to a double or single function item probe. The only other

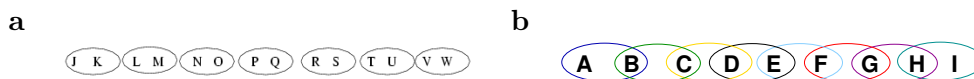


Figure 5. This example is representative of a list of words depicted by letters. Each item on the single function list is a novel word used in only one pair, while maintaining a solitary serial position relative to the pair. Double function items are presented in two separate pairs with differing serial positions within each pair. **a.** Single Function List. **b.** Double Function List.

possible circumstance that arises during the presentation of the cue is when the subject does not elicit any response to the stimulus. In this case, we simply characterize their error as *no response* and do not assimilate these errors with intrusive responses. The correct response to a cue, *C* (see Figure 5b), would be its corresponding pair member, *D*. A response of *B* would indicate a *backward* intrusion. Even though *B* and *C* do share group membership in a different pair, the order of presentation of the words in the pair is indicative of the correct response. The dual presentation of items in two separate pairs incites confusion, and often times lures participants to make incorrect responses. A *remote* intrusion to the cue *C*, could actually be one of several responses, including *E* or *A*, for example. Other than the *backward* intrusion and the correct response, a remote response embodies all of the words that are members of the same pair type as the cue, such as *I* or *F*. An *other* response can be thought of as being the functionally opposite response to a cue as the *remote* intrusion. Rather than being an inconsequential response to a cue from the same pair-type list, an *other* intrusion is any item that is a member of the opposite type of pair, for instance *J* or *R*, in response to the cue *C*. The final type of intrusion one could make is referred to as an *extra-list intrusion*, which is reasonably self-explanatory, in that the intruded item does not come from either of the lists, for example *Z*. The type of responses one could make to a single function probe follow similar guidelines to those constructed for a double function cue as in the examples above, but the unique properties of those responses will be described in greater detail below.

We broadly characterize all incorrect responses to a probe item as associative and non-associative intrusions. The intrusions falling under the category of associative intrusions include *backward* and *remote* responses. One theory attributes the formation of associations between items to overlapping temporal contexts of occurrence (Howard & Kahana, 2002; Howard et al., Submitted). A *backward* intrusion demonstrates this tendency clearly. For instance, *B* (see Figure 5b) was originally presented in two different pairs with overlapping contexts. A *backward* intrusion would presumably be a likely response to a double function item, due to its strong contextual and temporal overlap with the probe (e.g. *B* given *C*). *Remote* responses are also considered to be associative intrusions, even though these intrusions were never actually presented with the probe item in a pair. Since a temporal context is retrieved when prompted with *C*, items that share a context or reserve a strong overlapping context, indicated by a small linked-list lag, are more likely to be recalled than items that have a weaker contextual overlap with a large linked-list lag, or items with no overlap at all (Howard et al., Submitted).

Single function cues also induce *remote* and *backward* intrusions, but their connection to the probe is not so obvious. A sequence of single function pairs is arbitrarily constructed to form the single function list, and then the list is presented in a randomized order. Therefore, *backward* intrusions are not provoked by the sharing of common items during presentation. Instead, a *backward* intrusion indicates a lag of -1 from the cue item on the serial position list that was constructed for the purposes of making comparisons across lists. Because the single function pairs are not members of a linked-list like double function pairs, a *remote* intrusion does not involve any mediating items that facilitate direct connections across pairs. These intrusions are defined by any incorrect response from the single function list to a single function pair, excluding the *backward* response. *Remote* responses are potentially provoked by their similar membership with the same pair type.

Responses with no contextual or temporal overlap are considered to be non-

associative intrusions. These intrusions can be similarly explained for double function and single function cues. An *other* intrusion refers to any item that was originally presented as part of the oppositional list type (i.e. single function vs double function) of the cue. Another type of non-associative intrusion one could make is an *extra list intrusion*, which is indicative of a response that was not initially presented to the subject as part of the study list. *No response* by a participant is self-explanatory and implies a non-associative intrusion, but this type of response is not included in our analysis.

Methods. We conducted secondary analyses on a large paired-associate data set from a study reported in Howard et al. (Submitted). We divided the data by gender in order to examine its effect on the learning of double function and single function pairs. Gender was determined for the purpose of our analysis by referring to documentation collected at the time of the study. Two-hundred-six subjects, 147 women, were used for the gender analysis out of a total of 216 subjects who participated in the study. Three subjects' data were excluded from the analysis due to lack of paperwork indicating gender specification, and seven subjects (5 women) were cut because they did not properly complete the task, as indicated by their failure to meet the criterion of receiving a probability of correct recall > 0 for all trials of either pair type.

For a complete description of the methods used in this task, please refer to Howard et al. (Submitted).

Results. We measured the probability of correct recall and intrusions for both gender groups on each of the four trials. Comparisons were made between double function and single function pairs to examine learning across trials and frequency and type of intrusions. See Table 1 to view proportions of correct recall and intrusion types for single function and double function pairs across genders.

In order to consider the impact of gender on associative learning, we examined

Single Function													
Men							Women						
trial	cr	bk	rm	othr	xli	nr	cr	bk	rm	othr	xli	nr	
1	.299	.002	.043	.219	.096	.341	.369	.003	.037	.190	.068	.333	
2	.601	.002	.045	.149	.060	.143	.642	.001	.033	.128	.060	.136	
3	.716	.002	.018	.094	.072	.098	.748	.003	.020	.087	.050	.093	
4	.761	.000	.016	.082	.062	.080	.792	.005	.022	.069	.049	.063	

Double Function													
Men							Women						
trial	cr	bk	rm	othr	xli	nr	cr	bk	rm	othr	xli	nr	
1	.257	.158	.204	.043	.065	.273	.259	.180	.199	.035	.053	.273	
2	.386	.181	.149	.038	.060	.185	.415	.201	.159	.031	.044	.150	
3	.490	.166	.118	.027	.054	.145	.496	.188	.133	.023	.044	.116	
4	.540	.152	.107	.024	.052	.125	.546	.170	.121	.023	.039	.101	

Table 1: **Proportion of correct recalls and intrusions.** See text for a complete description of responses, cr: correct recall; bk: backward; rm: remote; othr: other; xli: extra-list; nr: no response

differences in probability of correct recall across trials and pair type. Figure 6 shows the proportion of correct recalls made by men and women for single function (Figure 6a) and double function (Figure 6b) pairs separately. It is clear from the picture that women outperform men on single function pairs, but they appear to have similar levels of performance when probed with a double function stimulus. We tested for significance by using a three-factor within-subjects ANOVA, with probability of correct recall set as the dependent variable, and trial, pair type, and gender set as independent variables. Consistent with previous findings (Howard et al., Submitted), a significant main effect of trial, $F(3, 1498) = 684.388$, $MSE = 11.061$, $p < .001$, pair type, $F(1, 1498) = 1049.848$, $MSE = 16.967$, $p < .001$, and an interaction of trial and pair type $F(3, 1498) = 36.590$, $MSE = .591$, $p < .001$ were observed. Interestingly, a new finding illustrated by the ANOVA revealed a significant interaction of pair type and gender $F(1, 1498) = 6.2982$, $MSE = .102$, $p < .05$.

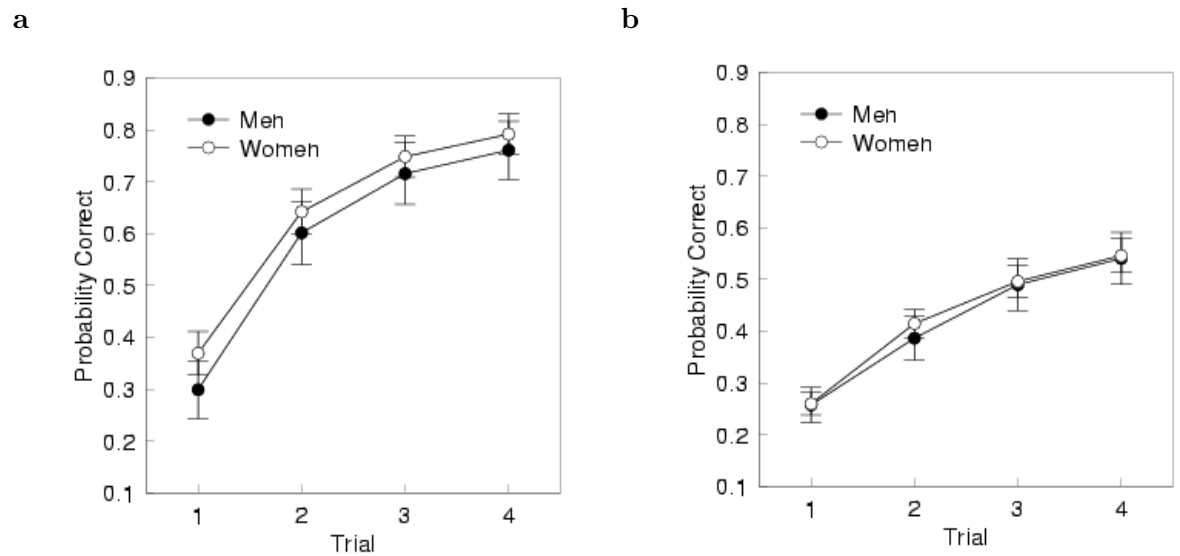


Figure 6. Probability of correct recall across trials for single and double function pairs. Error bars reflect 95% confidence intervals. **a.** Single Function. **b.** Double Function.

Associative Interference. The interaction of gender and pair type is evidence of women participants' superior ability to correctly learn single function pairs of words (Figure 6a), but not double function pairs (Figure 6b). Based on women's superior performance on single function pairs, we would expect the difference between men's and women's probability of correct recall to remain consistent across pair types. However, men and women maintained an approximately equivalent probability of recall on double function pairs across trials.

These results suggest that women are more susceptible to associative interference than men on double function pairs. Perhaps women were prohibited from achieving a higher probability of correct recall in relation to men, because their memories elicited other items on the double function list that were temporally and contextually associated with the probe. For instance, if women were more likely than men to remember a *backward* associated item to a cue, it may prevent them from delivering the correct response.

Given women's superior performance on single function pairs, we predict that

they are more susceptible to associative interference than men. If women do exhibit more associative responses than men to double function cues, this would reasonably explain why their probability of correct recall is comparatively lower than one would expect.

In order to investigate this possibility, we analyzed the intrusion data. We limited our intrusion analysis to incorrect responses that came from the study list, and ignored *extra list intrusions* and *no responses*. This strategy made it possible to assess specific transitions made between items on the list.

Figure 7 depicts the proportional relationships among the *backward*, *remote*, and *other* responses for single function (Figure 7a) and double function (Figure 7b) cues, given that the subject made an intrusion from the study list. The single function intrusion data seems jumbled and unsystematic, and without the presence of any apparent gender differences. Additionally, the intrusion data was limited for single function cues, because of the heightened probability of correct recall (see Table 1), resulting in minute values assigned to the remaining responses. The plots from the double function intrusion data are comparatively systematic and organized, with distinguishable gender differences. When an error is made from the list in response to a double function cue, women seem to be making more *backward* and *remote* responses than men. This can be seen by the systematic displacement downward of women's data plots, that is seemingly drawing them away from *other* intrusions, and pulling them closer to *backward* and *remote* responses.

To quantify Figure 7b, we tested for significance using a two-factor ANOVA. We constructed an associative index for each subject that compared the probability of making an associative intrusion versus a non-associative intrusion from the list (i.e. the probability of making a *backward* or *remote* response minus the probability of making an *other* response, divided by the combined probabilities of the three intrusions). We utilized this technique in order to control for inconsistent probabilities of recall and *no response* rates across genders for both pair types. Using the

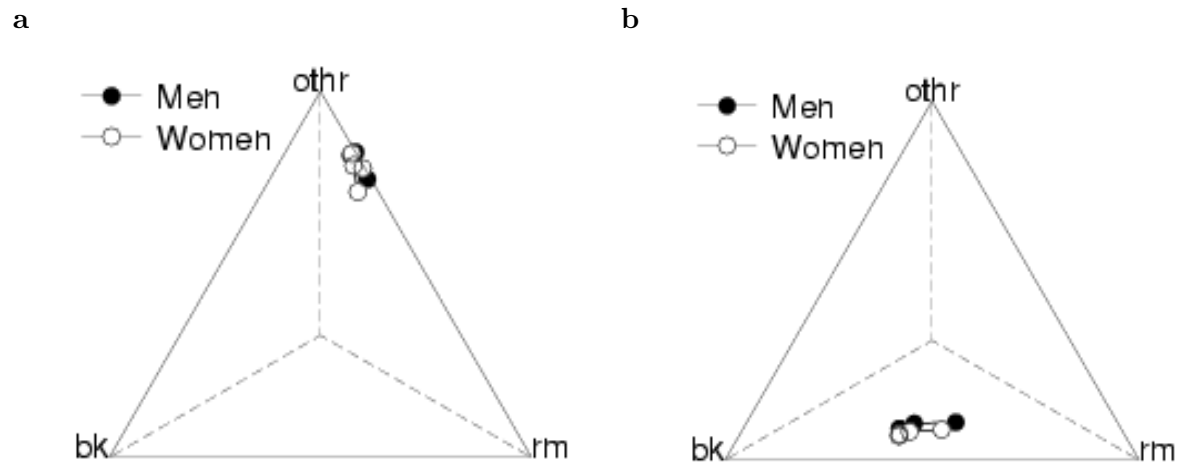


Figure 7. Proportion of *backward*, *remote*, and *other* intrusions in 3-D space, when all other types of responses are excluded. $Bk + Rm + Othr = 1$. **a.** Single Function. **b.** Double Function.

associative index as the dependent variable, and trial and gender as the independent factors, we observed a main effect of gender $F(1, 846) = 6.3125$, $MSE = .252$, $p < .05$ for double function cues, as we had predicted. The associative index from the single function intrusion data did not result in a significant effect of gender, $F(1, 480) = 1.9736$, $MSE = .796$, $p > 0.1$.

In response to a double function probe, women were significantly more likely than men to make an associative intrusion from the list rather than a non-associative response from the list. Even though men and women were seemingly performing at the same level on double function pairs according to their equal probability of recalls, women were able to better remember the context of the probe when they were unable to recall the correct answer. Presumably, women were more subject to the associative interference brought on by a double function cue, leading to a higher proportion of *backward* and *remote* intrusions in comparison to men. Somewhat counterintuitively, in this particular case, a better memory is reflected by excessive errors.

Recognition Memory

We initiated the examination of secondary analyses from two non-verbal tasks on the basis of gender in order to gain additional insight into the specificity of the gender differences that have previously been observed. By solely relying on the visual display of pictures to provoke learning, we compared recognition memory for men and women.

Recognition tasks differ from recall tasks, in that subjects are not expected to independently produce information from memory. Rather, recognition studies require participants to discriminate between items that have previously been studied and novel stimuli that were never before presented. A great deal of controversy exists in the field of cognitive psychology about the influences that determine recognition memory. Traditionally, dominant theories have attempted to describe recognition memory as being dependent upon a single process of familiarity, where responses are made on the basis of some threshold of confidence that subjects use to discriminate between “old” and “new” items (Yonelinas, 2001). However, in recent years more researchers have attributed episodic recognition memory to the contribution of two distinct components: recollection and familiarity (Yonelinas, 2001). Recollection can be thought of as a process synonymous to recall, where correct identification of an item relies upon the specific properties associated with the item and a concrete link to the original presentation of the item (Yonelinas, 2001; Schwartz et al., 2005). Other judgments made during a recognition study are thought to rely on the notion of familiarity, where responses are influenced by the similarity of the probe item and information previously stored in memory (Schwartz et al., 2005).

The rather intuitive approach of dual-process recognition models can be demonstrated in real-life situations. Suppose you reflect upon your commute to work last Friday morning. You may vividly recall walking in late at 10 : 00*a.m.*, because you missed the 9 : 15*a.m.* bus. During your stroll, you noticed the baby blue sky, illuminating sunshine, and cool breeze on this seasonally warm day in April,

that caused you to put on your sunglasses and carry your sweater. Your ability to recall your morning walk with a great deal of specificity about the time and context of the event would be classified as a recollected memory. On the other hand, imagine that you were asked to recall your commute to work last Friday, but a vivid description of your trip may not come to mind. You may make a few assumptions and generalizations about the transit to work that day. For instance, you know where you were employed at the time, you can vaguely remember missing the bus, but you do not recall how you actually got to work that day. Even though some general information is available after some rumination, you are unable to determine precisely the course of events that shaped your commute to work on that day. This type of generalization would be characterized by its familiarity-based nature. You are not able to restructure the event with specific information from that day, but you can make a reasonable estimate about the trip to work that day based on general knowledge and information you have maintained from that period of time.

We measure recognition performance by first classifying the types of responses one could make to a stimulus. In a recognition task that assesses memory of pictures, the subject must identify if (s)he had previously studied the item or if the picture was new to the task. If the cued picture was in fact an “old” item from the list, the subject could essentially make one of two responses. If the subject correctly identifies the item as “old”, the response would be considered a “hit”. However, if the subject mistakenly rejects the “old” item as “new”, the response would be categorized as a “miss”. Similar to the presentation of an “old” item during the test phase, if the subject is probed with a “new” item, there are two responses (s)he could issue. A “correct rejection” yields the appropriate answer, while a “false alarm” suggests that the participant misattributed the “new” for an “old” item. Because “hit” rate predicts “misses”, and “false alarm” rate specifies “correct rejections”, we primarily focus on these two types of responses when constructing descriptive figures for the data.

Using a dual-process approach, we attempt to measure recognition memory performance and classify decisions as either recollection or familiarity based judgments. The model is particularly useful when describing receiver operating characteristics (ROCs) and making comparisons across groups. A ROC graphs the proportion of correctly identified “old” items (hit rate) set as a function of the proportion of incorrectly recognized “new” items (false alarm rate) (Yonelinas, 2001). The graph is constructed by using a variable number of criteria set for the task. Rather than simply measuring “sure old” and “sure new” responses for each probe, recognition tasks provide a gradient of response criteria. This way, subjects are able to convey their level of confidence for each stimulus presented. The farthest left point on the ROC denotes the most stringent criteria, whereby the highest confidence responses are made. Each subsequent plot to the right of the first point include the cumulative values of less and less confident responses (Yonelinas, 2001) (see, for example, Figure 8). So, the hit rates and false alarm rates are relationally determined. In an experiment with six criteria, a response of 6 for an “old” item will always be a hit, while a response of 5 to an “old” item will be the cumulative proportion of hit rates to false alarms for responses 5 and 6. As the criterion becomes more and more liberal, the ROC approaches the maximum values of both probabilities (1, 1).

The Yonelinas dual-process high threshold model utilizes two free parameters to generate an ROC; discriminability d' and recollection R . Discriminability is computed as the average increase in familiarity assigned to items. This parameter measures the difference between the equal familiarity distributions of “new” and “old” items. Recollection is determined by the probability of recollecting an “old” item from the study list (Yonelinas, 2001). We use these two parameters to assess recollection and familiarity across genders using methods described in detail in Howard et al. (2006).

Methods. Analysis from two similar recognition tasks were assessed in order to determine gender differences in recollection and familiarity. This secondary analysis

was conducted after dividing the data based on gender identification specified at the time of the original studies.

Both studies used travel pictures as stimuli, taken from planetware.com, a travel picture website, and included a variety of scenes from throughout the world (Howard et al., 2006; Schwartz et al., 2005). The criteria to assess confidence levels was set to 6, so participants had six choices to use in response to one item. A response of 1 would indicate that the subject was sure the stimulus was a “new” item that had not been studied. Responding to an item with a 6 would indicate that the subject was sure it was an “old” item that had definitely been studied. Responding with one of the numbers in between the extreme values of 1 and 6 reflect less confident responses (Howard et al., 2006; Schwartz et al., 2005).

Experiment 1. The analysis from the first experiment included data from 19 women and 23 men. A detailed description of the experimental methods can be found in Howard et al. (2006).

Experiment 2. The secondary analysis from experiment 2 used data from 49 women and 21 men. A complete description of the experimental methods from this study can be seen in Schwartz et al. (2005).

Results. We used a dual process model to measure recognition memory performance and describe ROC curves. The estimated roles of recollection and familiarity can be observed by the curvature and shape of the ROCs. In the YHT (Yonelinas, 2001), if familiarity were the only process contributing to responses, ROCs are constructed that are symmetric along the diagonal line. If responses reflect some recollection based decisions, the leftmost point on the ROC will be shifted up, resulting in asymmetrical lines along the diagonal. Since the contribution of familiarity and recollection vary, the shape of the ROCs do not completely depend on levels of accuracy (Yonelinas, 2001). Although other approaches to fitting ROCs are possi-

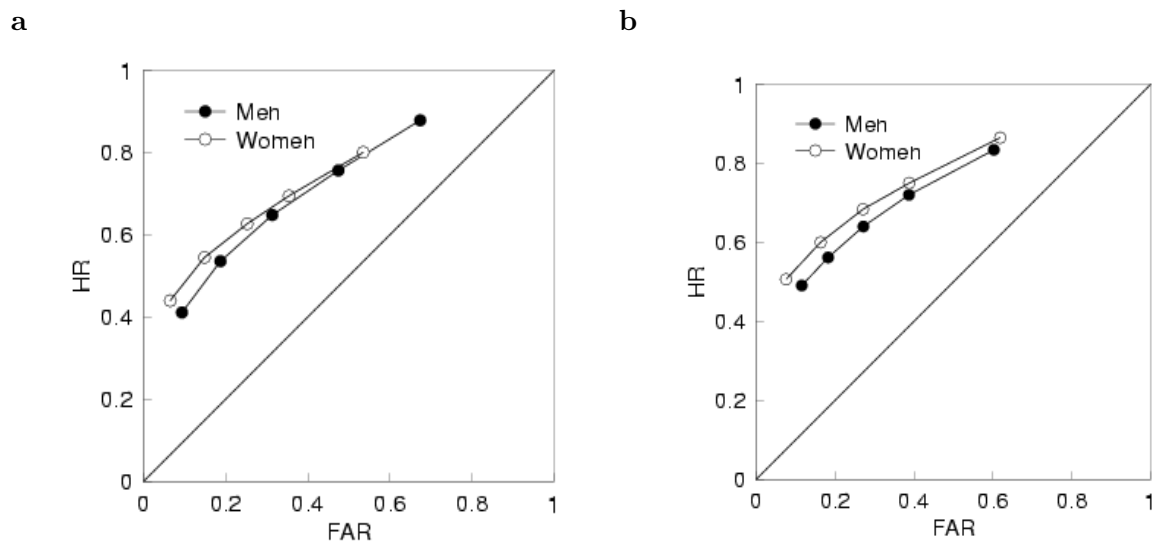


Figure 8. Receiver Operating Characteristic (ROC) curve, with hit rate (HR) set as a function of False Alarms (FA) shown for two recognition experiments. **a.** Experiment 1. **b.** Experiment 2.

ble, here the YHT provides superior fit to ROC curves from pictures (Howard et al., 2006).

The differences in familiarity and recollection contributing to men's and women's memory performance can be assessed by examining ROCs generated from experiments 1 and 2, see Figures 8a & 8b. Interestingly, the two independent components of recognition memory were not equally influenced by gender. The ROCs for men and women are similarly curved, but they elicit noticeable differences. Even though the shapes are relatively similar, the women's curves are clearly raised above the men's for high confident responses in both experiments. As the criterion becomes increasingly liberal, the distance between the lines is diminished. The graphs from the two recognition studies suggest that females have a higher probability of R , but levels of familiarity are approximately the same for men and women.

By applying the parameters provided by the model to our data, we were able to quantify the subtle differences observed from the ROCs. We compared recollection R and discriminability d' results from each task across gender groups by using t -tests.

Experiment 1 & 2 Model Parameters				
Exp	<i>R</i>		<i>d'</i>	
	Men	Women	Men	Women
1	.27(.03)	.35(.04)	.55(.08)	.55(.07)
2	.36(.03)	.41(.02)	.44(.06)	.55(.05)

Table 2: **Model Parameters for recognition experiments 1 & 2. The standard error of each mean is in parentheses.**

Assessment of the individual experiments separately did not yield a main effect of gender. This may be a consequence of the small subject pools allocated for each experiment, combined with the consistently small, but robust gender differences that have previously been observed in episodic memory tasks. In order to minimize the variability provoked by a small number of subjects in each gender group, we combined the results from experiments 1 and 2. The collective analysis was assessed by use of a two factor ANOVA with experiment and gender set as independent variables. Significant main effects of experiment $F(1, 109) = 6.603$, $MSE = .155$, $p < .05$ and gender $F(1, 109) = 5.469$, $MSE = .128$, $p < .05$ were observed when we considered *R* as the dependent variable. We did not observe significant differences in discriminability for gender or experiment.

Because the assessment of recognition memory is not necessarily straightforward, we have relied upon a model to describe performance across genders. We relied upon the parameters of the YHT dual-process model (Yonelinas, 2001) to quantify the gender differences from the recognition studies. We have used this model as a mechanism to substantively describe the gender differences we observed from the actual tasks. These findings from these two studies suggest that men and women rely equally on familiarity to make judgments. However, measurements of recollection, presumably a function that is dependent upon a specific memory of an event, are indicative of reliable gender differences favoring women.

Discussion

The intricacy of our reliable measurements have allowed us to the examine the impact of gender on episodic memory processes in great detail. Memory performance from the delayed free recall design did not differ across genders; however, a large gender effect was indicated by the math distractor task. Although discreet gender effects have been observed in some declarative memory tasks, we have neglected to identify an influence of gender that is of the same magnitude as the difference yielded by performance on the mathematical assessment.

The paired-associate task provided interesting findings regarding gender. Although women were able to correctly respond to a single function cue more often than men, double function probes did not elicit significant gender differences in probability of correct recall. Presumably, women did not surpass the performances of men on the double function cues because they were more susceptible to associative interference, indicated by their higher probability of responding with an associative intrusion to a double function probe in relation to men.

The two recognition studies, which assessed memory performance with the use of visually displayed travel pictures, elicited similar findings. The ROCs from the individual experiments clearly depict women's superior ability to recollect pictures, but familiarity for both genders is relatively indistinguishable.

Studies involving gender effects on cognition have typically relied upon two schools of thought to explain the source of the observed differences. The explanations occupy either the biological realm or they tend to favor social factors as being the primary source of influence. It seems likely that both social and biological factors play a role in generating differences between sex categories. Because direct connections between neurological modeling of the brain based on animal studies and cognitive models involving human behavior cannot be made, speculative implications can only be considered. Established biological sex differences cannot be simply translated into gendered cognitive differences, because these characteristics

are influenced by social and biological factors.

It is first essential to bear in mind several confounding factors that may have influenced the results from our secondary analysis. Unlike the delayed free recall task we designed and conducted, the data from the recognition and paired-associate studies was examined on the basis of gender after it was originally collected. Therefore, we were unable to control for several variables that typically arise during the course of the semester. For instance, the quality of our subjects varies throughout the semester, because the majority of our subject pool consists of students who are taking an introductory psychology class and volunteer to participate in our research studies for course credit. Those subjects who illustrate the tendency to procrastinate in fulfilling their research credit and participate toward the end of the semester usually do not provide the most useful data and may not match the caliber of the subjects who participated earlier in the semester. Although this is not necessarily the case, it could be confounded with gender.

Another factor to consider with all of the studies examined on the basis of gender is the limited subject pool we have access to for our experiments. All of our subjects attended Syracuse University, a private university in central New York, that mainly consists of white students who clearly have access to the privileged post-secondary educational system. Furthermore, recent trends indicate that the majority of people who attain advanced degrees in the U.S. are women (Peter & Horn, 2006). Given that more women sign up for our studies, declare Psychology as a major, and attend Syracuse University, it is reasonable to expect the women in our subject pool to be more motivated and focused while completing the experiment. If more women are demonstrating a genuine interest in attaining academic achievements than men, they may be putting forth a comparatively greater effort to complete the task. Although we did exclude those subjects who did not properly complete the task, we were unable to account for individual differences affecting focus, motivation, and attention.

Regarding the mathematical distractor task from the delayed free recall experiment, women's performances may be largely reflective of the stress induced by stereotype threat. Stereotype threat can alter behavior drastically, provoking personally harmful behavior associated with the internalization of particularly damaging stereotypes. Even though a woman may have the ability to do well on a math exam, she may perform below her potential due to the pressures and stress involved with the actual test, because of the negative constant reinforcement she has received throughout her entire life, from the media, education system, and the blatant reality of the workplace, where men overwhelmingly hold a majority of the careers in the mathematical and scientific domains (Fox, 2001).

To further speculate on the nature of the subtle gender differences we have observed in cognitive function, we will explore the influence of hormonal fluctuation in the female brain on memory. Here, we are not assuming that the participants who identified themselves as "female" in our cognitive studies had equal levels of estrogen, or that they necessarily maintained higher levels of estrogen than some subjects who indicated that they were "male". However, estrogen is an influential hormone in the female brain, and the potentially enhancing effects of estrogen on cognitive function provide reason to consider it as a variable for future research studies. Furthermore, it is particularly appealing to discover links between behavioral cognitive modeling and neurological findings in an attempt to determine the underlying causes of various observations across disciplines.

The hippocampus is the site in the brain where associations among items and events are formed and episodic memories temporarily stored (Bunsey & Eichenbaum, 1996). Many animal studies have investigated the impact of estrogen on hippocampal structures and processes. In vitro and behavioral studies have unveiled an array of enhancing effects evoked by the presence of estrogen in the hippocampus under natural and experimental conditions. Unfortunately, the beneficial properties estrogen apparently induces in the hippocampus of animals have not been directly

correlated to cognitive studies involving humans. However, due to estrogen's physiologically influential characteristics, it is of great interest to gather implications from neurological studies involving the impact of estrogen on brain structures believed to be responsible for basic cognitive functions.

Estrogen has been demonstrated to have an enhancing effect on synaptic plasticity in the hippocampus (Foy, 2001). Synapses are the communicative sites in the brain where messages are relayed between neurons. Synaptic plasticity refers to modifications to the formation or structure of the synapse following neuronal activation. Because of the elongated potency of the synapse following brief synaptic activity, researchers believe that synaptic plasticity plays an important role in memory storage. Cellular changes include an expansion of neural projections and a subsequent increased number of synaptic connections between neurons. These resultant alterations are thought to be closely associated with learning and memory, particularly when neurons within the hippocampus or the cerebral cortex are affected. Estrogen enhances this process by increasing the number of synaptic connections maintained by a neuron, and thereby potentially improving its ability to process information (Foy, 2001).

In conjunction with the augmented synaptic plasticity in the hippocampus provoked by the presence of estrogen, an increase in dendritic spine density on CA1 pyramidal cells is also observed. Dendritic spine density is influenced by the levels of estrogen present during the female rat's estrous cycle. When estrogen is at its highest level, during proestrus, spine density is similarly heightened. Concurrently, spine density reaches its lowest value during estrus when estrogen levels are diminished (Woolley & McEwen, 1994). The estrogen-induced increase in spine and synapse density has been positively correlated to working memory performance in mice (Daniel & Dohanich, 2001), where working memory can be classified in a synonymous category as short term episodic memory for humans. In a study using hippocampal slices of ovariectomized rats, following behavioral testing, Daniel

and Dohanich (2001) observed a parallel increase in NMDA receptor binding and working memory acquisition in an eight-arm radial maze task following the acute administration of estrogen 48 or 72 hours before testing.

The excitatory neurotransmitter, glutamate, mediates synaptic transmission within the hippocampus. Several different receptors bind to glutamate and cause the neuron to become excited within the hippocampus and provide input to CA1 pyramidal cells, including the NMDA (N-methyl D-aspartate) and AMPA (theta-amino-3-hydroxy-5-methyl-4-isoxazolepropionate) receptors. Estrogen is presumed to positively affect the growth of neural projections and enhance synaptic plasticity through both NMDA and AMPA receptors (Foy et al., 1999). Additionally, other studies have suggested that NMDA receptor activation is necessary in order to facilitate the heightened spine density of CA1 pyramidal cells provoked by estrogen (Woolley & McEwen, 1994).

The magnitude of excitatory postsynaptic potentials (EPSPs) within the hippocampus is intensified by the presence of estrogen. Particularly, AMPA and NMDA mediated response activity is heightened by estrogen (Foy et al., 1999). An *in vitro* study conducted by Foy et al. measured isolated AMPA and NMDA receptor-mediated EPSPs following the administration of estrogen to CA1 pyramidal cells located in the hippocampus. Estrogen rapidly enhanced both AMPA and NMDA receptor channels as evidenced by the increased amplitude of EPSPs evoked by Schaffer collateral stimulation (Foy et al., 1999).

Long-term potentiation (LTP) refers to an extended increase in synaptic transmission provoked by brief impulses of electrical stimulation in various neuronal pathways (Foy, 2001). Several studies have suggested that the sustenance of LTP may play an important role in learning and memory (Barnes, 2003). More precisely, LTP is currently viewed as a potential mechanism of memory storage within the hippocampus (Foy, 2001)

Within the hippocampal CA1 region, the induction of LTP requires stimula-

tion with a frequency that is high enough to sufficiently activate NMDA receptor pathways. Furthermore, the maintenance and expression of LTP is dependent upon enhanced AMPA receptor function (Foy, 2001). Estrogen has the ability to enhance synaptic transmission and LTP in CA1 neurons of adult, male rats, as indicated by the observation of hippocampal slices.(Foy et al., 1999). Additionally, studies involving awake, female, ovariectomized rats revealed an enhancement of LTP, either within 20 or up to 60 minutes, following the administration of estradiol benzoate to the CA1 region. Other in vivo studies have demonstrated the facilitation of LTP induction to be maximal in female rats during the afternoon of proestrus (Cordoba Montoya & Carrer, 1997).

The production of granule cells within the adult rat's dentate gyrus is notably boosted when estrogen is present. In addition to the granule cell proliferation induced by estrogen, the survival rate of these adult-generated cells is also enhanced. These findings have been implicated by studies involving the rat's natural fluctuation of ovarian hormones during their estrous cycle and in studies of animals that have been ovariectomized and injected with estrogen (Tanapat, Hastings, Reeves, & Gould, 1999).

Interestingly, female rats actually produce more new granule cells than males. This sex difference seems to be directly related to the heightened cell proliferation in the dentate gyrus of the female's brain during proestrus or immediately following the administration of estradiol. Production of new cells peaks during proestrus, although they rapidly deteriorate when estrogen levels are naturally minimized during estrus and correspondingly in cases where the rat has been ovariectomized. Subsequently, even though females generate more granule cells than males, the overall number of granule cells in the dentate gyrus is not significantly different across the estrous cycle. Specifically, a sex difference in the number of newly generated cells no longer exists between male rats and female ovariectomized rats that have received estrogen treatment 14 days prior to comparative testing. Moreover, a greater num-

ber of newly formed cells degenerate in the dentate gyrus of female rats than males (Tanapat et al., 1999).

The rapid formation of synapses among newly generated granule cells within the hippocampus has been implicated to play an important role in memory storage and retrieval. However, the impact of adult neurogenesis on memory and learning remains controversial. Studies have demonstrated that the survival of adult-generated granule neurons is enhanced in rats exposed to learning tasks that are dependent upon the hippocampus (Tanapat et al., 1999). Although the functional significance of adult neurogenesis is currently unknown, it seems plausible that it is essential to hippocampal integrity. Since it is the site in the brain where memories are stored and processed temporarily until they are transferred to other areas of the brain, the ongoing reproduction of new cells may be a likely contributor to the maintenance of these hippocampal functions (Tanapat et al., 1999).

Since the level of estrogen in the brain is essential to determine its influence on the neurophysiology of the hippocampus and its subsequent effect on cognition, cognitive studies measuring sex differences should assess hormone levels of participants at time of testing. The analysis we have provided does not account for the natural fluctuation of estrogen throughout the female menstrual cycle, since presumably the woman participants were tested at varying times of their cycle, and other factors that influence estrogen levels (i.e. birth control) were not considered. However, we can reasonably assume that the woman participants did have overall higher levels of estrogen than the male subjects. Regardless of the circumstances in our present study, estrogen levels and resultant hippocampal alterations provide prospective implications to the study of memory and future treatments available for cognitive deficits.

References

- Barnes, C. A. (2003). Long-term potentiation and the ageing brain. *Philosophical Transactions of the Royal Society B*, 358(1432), 765-72.
- Bunsey, M., & Eichenbaum, H. B. (1996). Conservation of hippocampal memory function in rats and humans. *Nature*, 379(6562), 255-257.
- Cordoba Montoya, D. A., & Carrer, H. F. (1997). Estrogen facilitates induction of long term potentiation in the hippocampus of awake rats. *Brain Research*, 778(2), 430-8.
- Daniel, J. M., & Dohanich, G. P. (2001). Acetylcholine mediates the estrogen-induced increase in nmda receptor binding in ca1 of the hippocampus and the associated improvement in working memory. *The Journal of Neuroscience*, 21(17), 6949-56.
- Fox, M. F. (2001). Women, science, and academia: graduate education and careers. *Gender & Society*, 15(5), 654-666.
- Foy, M. R. (2001). 17beta-estradiol: effect on ca1 hippocampal synaptic plasticity. *Neurobiology of Learning and Memory*, 76(3), 239-52.
- Foy, M. R., Xu, J., Xie, X., Brinton, R. D., Thompson, R. F., & Berger, T. W. (1999). 17beta-estradiol enhances NMDA receptor-mediated epsps and long-term potentiation. *The American Physiological Society*, 81(2), 925-9.
- Friendly, M., Franklin, P. E., Hoffman, D., & Rubin, D. C. (1982). The Toronto Word Pool: Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1,080 words. *Behavior Research Methods and Instrumentation*, 14, 375-399.
- Herlitz, A., Nilsson, L. G., & Backman, L. (1997). Gender differences in episodic memory. *Memory & Cognition*, 25(6), 801-11.
- Herlitz, A., & Yonker, J. E. (2002). Sex differences in episodic memory: the influence of intelligence. *Journal of Clinical and Experimental Neuropsychology*, 24(1), 107-14.
- Howard, M. W., Bessette-Symons, B., Zhang, Y., & Hoyer, W. J. (2006). *Aging selectively impairs recollection in recognition memory for pictures: Evidence from modeling and ROC curves.* (in press)

- Howard, M. W., Jing, B., & Probyn, J. P. (Submitted). Bridging the gap: Transitive associations are formed from learning items in similar temporal contexts. *Manuscript submitted for publication*.
- Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 923-41.
- Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269-299.
- Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle: teaching stereotype threat as a means of improving women's math performance. *Psychological Science*, *16*(3), 175-9.
- Kramer, J. H., Yaffe, K., Lengenfelder, J., & Delis, D. C. (2003). Age and gender interactions on verbal memory performance. *Journal of the International Neuropsychological Society*, *9*(1), 97-102.
- Lewin, C., & Herlitz, A. (2002). Sex differences in face recognition—women's faces make the difference. *Brain and Cognition*, *50*(1), 121-8.
- Peter, K., & Horn, L. (2006). Gender differences in participation and completion of undergraduate education and how they have changed over time. *Education Statistics Quarterly*, *7*(1 & 2).
- Primoff, E. (1938). Backward and forward associations as an organizing act in serial and in paired-associate learning. *Journal of Psychology*, *5*, 375-395.
- Pronin, E., Steele, C. M., & Ross, L. (2004). Identity bifurcation in response to stereotype threat: women and mathematics. *Journal of Experimental Social Psychology*, *40*, 152-168.
- Schwartz, G., Howard, M. W., Jing, B., & Kahana, M. J. (2005). Shadows of the past: temporal retrieval effects in recognition memory. *Psychological Science*, *16*(11), 898-904.
- Tanapat, P., Hastings, N. B., Reeves, A. J., & Gould, E. (1999). Estrogen stimulates a transient increase in the number of new neurons in the dentate gyrus of the adult female rat. *The Journal of Neuroscience*, *19*(14), 5792-801.

- Woolley, C. S., & McEwen, B. S. (1994). Estradiol regulates hippocampal dendritic spine density via an N-methyl-D-aspartate receptor-dependent mechanism. *The Journal of Neuroscience*, *14*(12), 7680-7.
- Yonelinas, A. P. (2001). Components of episodic memory: the contribution of recollection and familiarity. *Philosophical Transactions of the Royal Society B*, *356*(1413), 1363-74.