Department of Psychology, University of California, Los Angeles

Fall 2014

Table of Contents

Analogical Reasoning and Problem Solving .......................................................... 3-10
Aditi Gala, Faculty Mentor: Hongjing Lu, Ph.D.

Review of an Emotional Intelligence Study ......................................................... 11-14
Ben Bordofsky, Faculty Mentor: Elizabeth Bjork, Ph.D.

Error Generation, Feedback, and Curiosity ....................................................... 15-21
Charlie Hall, Faculty Mentor: Elizabeth Bjork, Ph.D.

Functional Connectivity Analysis ................................................................. 22-27
Charlotte Cornali, Faculty Mentor: Jesse Rissman, Ph.D.

Flash-Lag Effect and Inverted Biological Motion ............................................. 28-33
Chin-Wei Andy Huang, Faculty Mentor: Hongjing Lu, Ph.D.

Nature Letter Assignment ................................................................. 34-38
Cody Sakurao, Faculty Mentor: Zili Liu, Ph.D.

Effects of Interactivity on Perceptual Learning in Mathematics ....................... 39-47
Emma Murray, Faculty Mentor: Phil Kellman, Ph.D.

Virtual Reality, Context Dependency, and Learning ........................................ 48-53
Jacob Yu Villa, Faculty Mentor: Jesse Rissman, Ph.D.

Human Perception Laboratory ................................................................. 54-61
Jason Carpenter, Faculty Mentor: Phil Kellman, Ph.D.

The Study on Self-Regulated Learning through Learning Painting Style .......... 62-68
Jingqi Yu, Faculty Mentor: Elizabeth Bjork, Ph.D.

The Influences that Guide Judgements of Learning ......................................... 69-74
Joselyn Ho, Faculty Mentor: Alan Castel, Ph.D.

Individual Differences in Perceptual Learning: Potential Basis for Differences in General Intelligence ................................................................. 75-81
Kyleigh Smith, Faculty Mentor: Phil Kellman, Ph.D.
The Effect of Movement Control and Navigation Control on Spatial Knowledge Tasks... 82-88
Kylie Springsteen Faculty Mentor: Zili Liu, Ph.D.

Effects of Interactivity, Sequencing, and Retirement on Factual Learning ................. 89-110
Lorena Hinckley, Faculty Mentor: Phil Kellman, Ph.D.

Neuropsychological Disorders and Memory ......................................................... 111-116
Mariam Hovhannisyan, Faculty Mentor: Barbara Knowlton, Ph.D.

Mind Wandering Attention .................................................................................. 117-122
Michael Ruder, Faculty Mentor: Jesse Rissman, Ph.D.

Social Working Memory ....................................................................................... 123-129
Natalie Saragosa-Harris, Faculty Mentor: Matthew Lieberman, Ph.D.

Discovery and Fluency Effects in Adaptive Perceptual Learning ......................... 130-137
Nicole Mirea, Faculty Mentor: Phil Kellman, Ph.D.

Mean Diffusivity and the Microstructural Properties Underlying the Hippocampus... 138-143
Omri Raccah, Faculty Mentor: Jesse Rissman, Ph.D.

Learning and Forgetting Lab: Confidence Weighted Multiple Choice Testing ......... 144-149
Simran Grewal, Faculty Mentor: Elizabeth Bjork, Ph.D.

Using Photographic Life Logging to Explore Autobiographic Memory ................. 150-155
Tim Yu, Faculty Mentor: Jesse Rissman, Ph.D.

Effects of Form of Stimulus on Analogical Problem Solving ............................... 156-164
Yao (Sherry) Hsiao, Faculty Mentor: Hongjing Lu, Ph.D.

**These research papers are written by undergraduate students as part of the capstone requirement for the Cognitive Science major.**
Analogical Reasoning and Problem Solving

Aditi Gala

The ability to reason by analogy underlies human intelligence and allows us to make sense of new information by relating it to prior knowledge, and also gives us the ability to look at familiar information with strange, new perspectives. This ability can be very useful in constructing new models, solving novel problems, making predictions, designing experiments, and constructing arguments (Gick & Holyoak, 1983). Analogical reasoning requires the understanding of abstract concepts or schemas that can be applied across systems that are superficially semantically disparate. This transfer of knowledge involves the process of mapping relevant aspects of the source system (a system that we are already familiar with, say for example, the solar system) to corresponding aspects of a target system (an unfamiliar system that we are trying to understand, say for example, the structure of an atom). Gick & Holyoak (1983) provide a theoretical analysis of the mechanisms of the mapping process. The transfer of information could happen at the same level of abstraction by comparing concrete analogs from different systems, like the heart and a water pump. Another similar mapping process is involved in comparing a concrete analog to an abstract schema, example the heart and the general concept of a "pump". Furthermore, mapping is also involved in abstracting a schema from two concrete analogs, example understanding the idea of a "pump" by comparing the heart and a water pump. The schema abstraction happens by "eliminative induction" which entails removing the differences while preserving the similarities in causality.

There are many factors that influence the ability to successfully reason with analogy. The most important of these is the level of understanding of the source system. A deeper understanding of the source analog corresponds to a more concrete higher order mental
representation of the underlying elements, relationships and patterns. It is known that the presentation format of the material has an influence on the acquisition of knowledge. The majority of prior research on analogical reasoning, however, only involves verbal representations of the source system. Although there has been some research done on the benefits of using animations to depict source systems (Pedone, Hummel & Holyoak, 2001), analogical transfer from pictoral representations, such as diagrams or animations, as well as from multimedia representations needs further investigation. Another factor that plays an important role in analogical reasoning is the ability to spontaneously notice the relevance between the source system and the target problem. Most prior studies have found that people often have trouble spontaneously noticing this relevance. Providing a hint about the underlying similarity to the source analog significantly improves participants ability to solve the target problem. However, one study in particular found seemingly contradictory evidence (Day & Goldstone, 2011). They found that participants often generated implicit mappings without even realizing it, and furthermore being told to form explicit correspondences between the source and target systems actually impaired performance. The different nature of the source system and its presentation in this study could be a possible explanation for these findings.

This study aimed to investigate the role of animated, multimedia presentation on analogical reasoning and problem solving. A source system was designed that involved cannons and a friendly barrier surrounding an enemy. It was hypothesized that supplementing the verbal instruction of the source system with pictoral representations will improve the subjects' performance on the target problem because using two information encoding channels simultaneously will result in a better, richer mental representation of the relevant causal relationships in the system. Furthermore, it was also predicted that multimedia representation of
the source system will result in more participants spontaneously noticing its relevance to the target problem because of an increase in the probability that the source system will be retrieved.

Method

Participants

A total of 126 subjects participated in this study. They were all UCLA undergraduate psychology students who were compensated with course credit for participating in the study. The subjects were equally divided into three conditions: Verbal Animation, Verbal Diagram and Verbal. Subjects that were familiar with the target problem or the concept of analogical reasoning in general were excluded from the study.

Materials

The instructions, questions and stimuli were presented on a CRT monitor at a resolution of 1024 x 768 pixels. The animations were prepared using MATLAB and Psychtoolbox. An abridged twelve-item version of Raven's Progressive Matrices test was used as a filler task as well as a measure for general intelligence. Qualtrics was used to administer the questions in a survey format.

Design and Procedure

Participants were randomly placed into one of the three conditions. The same four scenarios were presented in each condition. Instructions explaining the source system's objective, elements and their relations preceded each of these scenarios. In the Verbal Animation condition, a spoken monologue along with a supplemental animation was presented for each scenario. The Verbal Diagram condition the spoken monologue was accompanied by static
diagrams. In the Verbal condition the subject is presented only with the spoken monologue. The general schema being instructed was the convergence principle. The source system involved either one or more cannons attempting to destroy an enemy without hurting the surrounding friendly barrier. While the small cannonballs did not damage the friendly barrier, individually they were too weak to damage the enemy. On the other hand the large cannonballs did succeed to damage the enemy but not without causing significant harm to the barrier first. The radiation problem was used as the target problem to assess analogical transfer. Scenario #4 is analogous to the solution of the radiation problem as it represents the convergence principle being applied successfully.

Following each of the scenarios the participants were presented with a free response asking why the cannons failed or succeeded in that particular scenario. After going through all the scenarios the participants were asked to complete the abridged version of the Raven's Progressive Matrices. After the completion of the task, participants are presented with the radiation problem without any explicit hint indicating that it is related to the previously seen scenarios. After the responses are submitted and recorded the participants are once again asked to provide a solution to the radiation problem but this time they are given a hint to recall the four scenarios. Finally, the subjects are asked if they noticed the relevance of the scenarios to the radiation problem before the hint and if they had encountered the radiation problem prior to the experiment.

The transfer rates were scored as the following: (1) failed to solve the radiation problem, (2) solved the radiation problem successfully with the hint, (3) spontaneously solved the radiation problem (without the hint). The free responses to each of the scenarios as well as the radiation problem were assessed using the same grading criteria by two different researchers.
The assessment of a third researcher was used to settle any disparities in the scoring by the two original graders.

Results

The subjects' ability to spontaneously or directly transfer knowledge from the scenarios to the analogous radiation problem was scored depending on whether they conveyed at least two of three ideas critical to the convergence principle: *multiple* radiation sources (1) positioned *around* the tumour (2) firing *low intensity* rays (3). When analysing the transfer rates to the radiation problem, a chi squared test was used. Figure 1 shows the spontaneous as well as total transfer rates for each of the three conditions. In agreement with the prediction that was based on existing literature, the radiation problem spontaneous transfer rate in the Verbal Animation condition was significantly better than those found in both the Verbal Diagram \( \chi^2(1, N = 84) = 10.50, p = .001 \) and Verbal conditions \( \chi^2(1, N = 84) = 8.02, p = .005 \). Interestingly, however, the total transfer rates of the radiation problem in the Verbal Animation condition was slightly but not significantly better than the Verbal condition \( \chi^2(1, N = 84) < 1 \) and was significantly higher than the transfer rates for the Verbal Diagram condition \( \chi^2(1, N = 84) = 4.53, p = .03 \).

These results indicate that, contrary to expectation, the supplementary diagrams do not improve performance on analogical transfer rates, spontaneous or otherwise. However, if the supplementary pictoral information is animated it does improve the subject's ability to spontaneously notice the relevance between the source system and the radiation problem.
Figure 1: Spontaneous Radiation problem transfer rates (left) and total transfer rates (right) for Verbal, Verbal Diagram and Verbal Animation conditions.

To assess the subjects understanding of the source material, the scenario free response scores and multiple choice scores were compared using an independent samples t-test. The analysis shows a main effect of condition. The Verbal Diagram condition had significantly higher scores than The Verbal only condition \( t(82) = 2.23, \ p = .03 \). The Verbal Animation condition scored significantly better than both the Verbal Diagram \( t(82) = 2.27, \ p = .03 \) and Verbal only condition \( t(82) = 27.00, \ p < .001 \). From these results we can conclude that supplemental pictoral information leads to a deeper understanding of the source material. The results agree with what existing literature suggests, supplemental pictoral information in the form of animations is even more effective in the understanding of source scenarios than diagrams are.
Figure 2: Source understanding scores for Verbal, Verbal Diagram and Verbal Animation conditions. Free response (left) and multiple choice (right).

Discussion

The increasing use of multimedia instruction in educational institutions today makes it important to study the nature of analogical reasoning. The results from this experiment show that supplementing the verbal presentation with an animated pictorial representation not only benefits the participants understanding of the source system but also helps with spontaneously noticing the relevance between the target problem and the source analog. This could be because the animation provides the participants with a more robust abstract representation of the relevant causal relationships. Another possibility is that the brain regions employed when encoding and processing dynamic visual information are simply more conducive to better retrieval. When the supplemental pictorial information is static in the form of diagrams, the participant's understanding of the sources system did improve, however it surprisingly seemed to worsen their ability to transfer the relevant information to the target problem. This detrimental effect of using
diagrams in analogical problem solving may be limited to systems and schemas that are dynamic (i.e. varying with time) like the radiation problem and the convergence principle; however, further investigation is needed.

The hint provided in the experiment resulted in a significant improvement in the participants ability to solve the target problem. This shows the importance of having a "teacher" to point out the relevance of two superficially dissimilar systems and gaining an understanding of the abstract schema that is common to both. The findings of this study indicate that incorporating animations in educational instruction would promote learning and more importantly self learning. On the other hand, diagrams for the instruction of class material should be used with caution.

References


In this paper I will be reviewing the research performed by Curci, Lanciano, and Soleti titled, *Emotions in the Classroom: The Role of Teachers’ Emotional Intelligence Ability in Predicting Students’ Achievement*. The basic premise of their research is that the emotional intelligence of the teacher can affect student’s performance both positively and negatively.

Emotional Intelligence (EI) when high in the teacher can promote students’ self-perceptions of ability and self-esteem which have both been found to be indicators of students’ success. There are two primary ways to view EI. The first is as a set of personality traits. That is to say that someone’s personality can be more or less emotionally intelligent. The other way to view EI is as an ability. That is to say that people can control and use it in different ways. The researchers in this paper chose to study EI from the ability perspective. This was a good choice because it allows for the researchers to test something that is more useful. EI as personality traits is not as useful as ability because personality traits are often blended and have much overlap thus rendering them difficult to distinguish and test for. Additionally, personality traits are not something that can be tangibly changed or modified very easily. So even if the researchers were to find that teachers with high EI as personality traits had students with higher test scores, it would have fewer implications than EI as an ability. If EI as an ability is found to be significant in affecting students’ performance, the implications for improving teaching methods are significant.

The researchers frame their question in the following manner: “to what extent [does] teachers’ level of EI ability…moderate the impact of students’ self-esteem, perception of
ability, and metacognitive beliefs on achievement in mathematics and sciences?" In short, does teachers’ EI ability affect student performance? Following this question, the researchers present four hypotheses. First, teachers’ EI abilities are related to their emotional states and self-efficacy.

Second, student achievement is related to their self-esteem, self-reports of ability, and metacognitive beliefs. Third, teachers’ EI abilities have a predictive effect on student achievement. Lastly, students’ self-esteem, abilities, and metacognitive beliefs interact with teachers’ EI abilities. By presenting four hypotheses the researchers cover the most important aspects of the question they are asking. The question they are asking is too complex for a single hypothesis.

Testing EI, as well as self-esteem, metacognitive beliefs, emotional states, and self-efficacy is very challenging. The primary way in which to test these factors is through a battery of questionnaires and surveys. This brings up the common issues of self-reporting reliability/accuracy and the external validity of the questionnaires and surveys. However, the researchers in this study do a fair job of recognizing and combating these issues. First, they tested each factor separately using a battery of tests that have been used successfully in other research. By testing each factor separately they are focusing their measures which allows them to mitigate any factor overlap and to eliminate some of the confounding factors. Using tests and questionnaires that have been tried and tested in other research adds to the validity of the test results.

An important measure that should be discussed in more depth is the measure of EI. The researchers for this study used an Italian version of the MSCEIT. This particular test is a series of over one-hundred multiple choice questions which are then scored by consensus. By using
consensus scoring the researchers were able to increase the validity of this test compared to using a single judge or a formulaic scoring method. There are of course limitations of using human’s to score a test, but given the difficulty in measuring EI to begin with, the measures taken to combat this seem reasonable and reliable.

One measure that may be questionable was the use of GPA as a measure of achievement. This measure does not have much variance and could therefore be hiding some significant results. A final test, or a series of tests throughout the school year may have been a better measurement to show achievement and progress in the student’s work.

Another potential issue with the researcher was sample size. They had an ample amount of students (N=358) but only twelve teachers. However, this issue can be assuaged by a careful presentation of how generalizable the study is. The researchers also took this into account during their statistical analysis. They presented research that has shown that a multilevel model can still be effective even with a sample size of only three.

This research used a multilevel modeling approach because the data collected is hierarchical and multi-dimensional. After running a series of complex statistical analysis the researchers found only one significant result. That result being that out of all the teacher characteristics measured, only emotional intelligence abilities were the only ones that positively correlated with an increase in the students’ performance. This was as the researchers predicted. Given that we know self-esteem and self-reported abilities correlate with student success and teachers’ EI correlated with student performance it follows that teachers’ EI interacted with the students’ self-esteem and self-reported abilities. This increase in self-esteem and self-reported abilities can then be seen as an explanation for the increase in the students’ math/science GPA. While it is disappointing that none of the other factors that the researchers tested were found to
interact significantly with student performance, this is not particularly surprising. While the researchers took the necessary steps to try and distinguish each factor it is still incredibly difficult to single out personality traits, especially when self-reported. Improvement in the ability to measure these factors could lead to further significant results.

Another issue that the researchers mention briefly is the gender of the teacher. The way students relate to a male versus a female teacher could be a confounding factor. In order to combat this further research should examine equal numbers of male and female teachers.

Another related issue not mentioned by the researchers is culture. All subjects in this study were Italian. It is not unreasonable that the ways in which people relate and interact in Italian culture varies significantly from other cultures. This is another limiting factor on how much this study can be generalized. Further research is needed with different cultures.

While the generalizability of this study is constrained, the findings could lead to further research that changes how we train teachers. The implications of the findings of this research are very important. Since EI is considered an ability this means that it can be taught, trained and improved. EI training could be integrated into all teaching credential programs and improve education vastly. This knowledge could also help struggling students; rather than targeting performance directly, teachers could use their EI abilities to improve students’ self-esteem which in turn will help performance. This could be an effective strategy. Use of EI ability tests could also be used in the hiring process of teachers. It is important that teachers’ skills are recognized and that they are rewarded for them. The implications are many and have strong potential to benefit teachers and students. Further research needs to be done in this field so that we can confirm and expand these findings.
Error Generation, Feedback, and Curiosity

Charles D. Hall

Introduction
Retrieving information from memory is essential to maintaining long-term mastery over the material. The benefit of this retrieval practice, known as the testing effect, is known to exist even when the initial retrieval is incorrect. One possible explanation for the benefit of pre-testing suggests that an increase in interest in the topic, or curiosity, may heighten the anticipation for the correct answer, subsequently causing the correct answer to be more memorable. The goal of this paper will be to discuss the background of what is known about pre-testing, feedback, and curiosity as well as discuss a possible study that could tease out how differences in levels of curiosity might affect error generation effects.

The Error Generation Effect

Given that many studies have shown that generating a response strengthens the probability that one will remember that response again at a later time, the error generation effect postulates that the act of generating an answer, regardless of whether it is correct or not, increases long-term retention. One study done by Kornell et al designed six experiments to test the differences between error generation and read-only study conditions. In order to ensure that the study focused on irretrievable test items, the experiments featured both plausible fictional trivia questions (“What is the last name of the person who panicked America with his book Plague of Fear?”) and weak associative pairs (whale –mammal). For the fictional, yet plausible trivia questions, the testing and read-only trials consisted of both fictional and nonfictional trivia which served two purposes. First, it tricked participants to view the fictional trivia questions as real questions for which they could not retrieve the answer rather
than completely fictional questions. Secondly, by comparing the generation effects of the fictional answers, most incorrectly guessed at first, and the nonfiction answers, correctly guessed at first, the results could more clearly pinpoint the effects of incorrect generation independently from the effects of testing and/or generation itself. In both the experiments involving fictional trivia and word pairs, study time was manipulated so that learners in the read-only conditions would see the paired word for at least the same, often more, time than the error-generating learners. This assured that the results indicating that error generating effects were more beneficial than read-only could not be attributed to study time. In the end, all manipulations of study time, final test delay, and within and between-subject design led to greater long-term retention for questions in which subjects generated incorrect responses before learning the correct answer.

Feedback

If unsuccessful or incorrect retrieval aides long-term retrieval, how might varying the ways in which feedback is provided alter this effect? Feedback can come in many forms including immediate, delayed, and variable. In 2014, Kornell again examined the effects of pre-testing, this time by examining the role that delayed feedback had on information. In his first experiment, memory for word pairs (e.g. antler-horn) were shown to only benefit from immediate feedback; however, when the stimuli were trivia questions as in Experiment 2, delayed feedback was also found to lead to better retention than study-only conditions. While these findings do not suggest that delayed feedback is more beneficial than immediate feedback, they do raise important questions about the mechanisms at work when one waits
One possible explanation for the benefit of delayed feedback may relate to the level of curiosity one has for learning the correct answer. Mullaney et al. examined how varying levels of curiosity might interact with varying type of feedback including immediate, delayed, and varied delay conditions. In this study, participants were asked to give their best guess for seventy trivia questions and were subsequently asked to rate their curiosity to know the answer on a scale between 1 and 6. After giving their curiosity rating, subjects were either shown the correct answer immediately, or after some delay. In Experiment 2, they included a varied delay condition that showed subjects the correct answer after 2, 4, or 8 seconds. The results indicated that participants were more likely to recall answers to questions they were more curious about. Furthermore, for questions rated low in curiosity, no effects of immediate versus delayed feedback existed; however, when questions were rated high in curiosity, participants benefitted from varied delay. This suggests that delayed feedback may be beneficial when participants are more aroused and therefore paying more attention. In her paper, Kellie Mullaney offers a potential explanation for why delayed feedback benefits are present for high curiosity questions, “this process is less likely to occur when the feedback is immediate, because the instantaneous presence of the answer cuts short any opportunity to engage in this anticipatory process”.

Furthermore, Mullaney offers one important possibility for what type of processing occurs during the feedback delay:
One factor influencing curiosity could be some degree of partial knowledge that participants have about the answer to a given question. For example, even if they could not recall what a deltiologist collects, it is possible that participants have some familiarity with this word (perhaps having heard it before but not knowing what it means), and this sense of familiarity could drive their curiosity to see the answer. Curiosity judgments could also reflect confidence: Participants may be more curious about answers that they feel they have gotten correct or that they feel they should know.

This explanation seems to suggest that items that are rated higher in curiosity may be linked to a feeling that one “could have” come up with the correct response. While this explanation may justify why delayed feedback benefits questions rated higher in curiosity, no research has combined measures of curiosity, specifically “could have” known judgments, with the effects of pre-testing versus study-only.

**Potential Study**

Since curiosity seems to play a role in the differences observed in delayed versus immediate feedback, varying levels of curiosity may play an important role in the benefits of pre-testing versus read-only conditions. A potential study would examine the differences observed between generating a response before seeing the answer and read-only conditions in relation to how curious participants were to know the answer. Curiosity would be measured by asking participants to rate, on a scale from 1 to 6, how much they felt they “could have known” the answer. This rating would be made immediately followed by the 12 second study phase detailed below. The first row represents the generation trials that asks the participant to generate a response before they view the correct answer. The second row represents the read only condition. In both the generation and read-only trials, the question will remain on the
screen for the same amount of time. Notably, the correct answer will only appear in the generation trials for four of the total twelve seconds.

<table>
<thead>
<tr>
<th>8 seconds</th>
<th>4 Seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td>What year did the Golden Gate Bridge open?</td>
<td>What year did the Golden Gate Bridge open? 1937</td>
</tr>
<tr>
<td>____</td>
<td>1937</td>
</tr>
</tbody>
</table>

Immediately following the above presentation, participants will be asked to evaluate how well they “could’ve known” the answer on a scale from 1 to 6. This response will correlate to how much participants are familiar with the subject matter of the question. Optionally, a judgment of learning measure, JOL, may also find out how confident participants are in later recall of the answer. A JOL may be beneficial because it may reveal differences between high-curiosity items and high-confidence items.

**Expected Results**

Overall, I expect generation trials to lead to better recall than read-only trials. Further, when a participant provides a measure of how much they “could have known” a trivia question, this measure should reflect whether the participant just experienced not being able to generate the answer. That is to say, more questions in generation trials may end up with lower “could have known” judgments and more read-only trials may
be higher in “could have known” judgments. This may be the result of hindsight bias that participants think they could’ve gotten the question right if they were asked. In terms of recall on the final test, I predict that for questions in generation trials, higher “could have known” judgments will lead to higher recall performance. Conversely, for questions in the read-only trials, recall may be higher in lower “could have known” trials. These predictions rely on the articles described above that suggest that higher anticipation leads to more attention to the answer. In trials where generation was required, more familiarity might cause a feeling that not only “could” the participant have known the answer, but in fact they “should” have known the answer. In read-only cases, overconfidence may cause participants to attend more to answers they do not feel they could have known and pay less attention to answers they feel they could have known.
Works Cited


Due to the significant advancements in technology, today, there are more ways to examine the inner workings of the human brain. The presence of spontaneous blood-oxygen-level-dependent (BOLD) signals of functional magnetic resonance imaging (fMRI) has been used to assess functional connectivity. Consequently, the research produced is aimed at further examining resting state networks (RSNs) and how the structural and functional connectivity are related. It was previously believed that during rest, the brain connectivity was stationary, however, it appears that this is not correct. New research has demonstrated that the connectivity between different nodes is actually non-stationary at rest. This connectivity can be represented in the form of a graph or network of interconnected nodes, linked by a set of edged nodes (Monti et al., 2014).

The dynamical organization of the nodes is related to the underlying structural connections, thus, functional and structural connectivity should be analyzed concurrently. Vuksanović & Hövel (2014) have suggested that there are resting-state functional networks, regions activated at rest with no explicit behavior, and observed functional correlations between cortical regions without any direct neural connections. The BOLD signal fluctuations are highly structured and the robust patterns of correlated activity arise from the underlying neural dynamics and structural connections still need to be more thoroughly examined (Vuksanović & Hövel, 2014).

The majority of the current research is focused on the functional connectivity correlations between regions of interest (ROIs) in order to better comprehend the dynamic evolution of
networks which may lead to gain a better understanding of the brain’s organization and
cognition, which are still very much a mystery. Current research project hope to identify if there
are functional connectivity profiles that have predictive power over behavioral measures, such as
IQ. There have been numerous studies trying to examine the relationship between varies
functional brain networks that exhibit high degrees of intrinsic functional coherence and how the
number connective fibers relates to those differences (Vahktin et al., 2014, p.349). Vahktin et al.
(2014) used the Ravens Progressive Matrices to examine fluid intelligence. This test represents
relatively complex cognitive processes. They hypothesized there would be widespread spatial
redistributions of functional networks during the RPM test compared to rest (Vahktin et al.,
2014, p.350)

I have been assisting Nicco Regente, my sponsoring Graduate Student, with his work for
Professor Rissman in his neuroscience research using the Human Connectome Project. Since the
structural connectivity analysis was already complete, I was tasked with starting the functional
connectivity analysis. The aim of this project is to assess if there is any correlation between a
person’s functional connectivity profile and behavioral measures, in the hope of discovering if
functional connectivity can have some predictive value of behavioral measures.

During the quarter, I have been writing MATLAB scripts to collect the necessary resting
state data and organize them according to each subject. Further analysis will be conducted on
the data from resting state fMRI images collected from the Human Connectome Project third
quarter data. The scripts pulled from all the subject folders of the Human Connectome Project
folder and copied the folders onto the FUNC.
Using MATLAB’s imaging (SPM) software, I prepared each subject’s resting state file, rfMRI_REST1_LR, by splitting them into individual images for the functional connectivity analysis. However, since each subject file was composed of a 4D structural image and SPM can only process 3D images, I had to split the 4D files into separate 3D images, so that each subject had as many images as time points. This was accomplished by using MATLAB’s SPM function spm__split_vol(). The reason for splitting the images was to create a time course of activity, a 4D array composed of 3D images, so that the functional connectivity analysis could be conducted. After preparing the files, I began writing another script to create a “mean time course” for all the voxels in each of the 264 ROIs of interest at each time point. Since we are already using predefined ROIs, there was no need to write scripts to identify the independent components, or nodes, using Independent Component Analysis. Collecting the data for each mean time course from the voxels within each ROI will provide data needed to begin the functional connectivity analysis.

Integrating the results from both the structural and functional connectivity analysis is necessary for developing an understanding of the structural basis of functional connectivity, in hopes of exploring the possibility that connectivity profiles could have predictive implication of certain behavioral measures; this type of integrated research will be important. However, there are limitations for future research that directly compares the two, which was made evident in Khalsa et al. (2014). They examined the structure-function relationship of the posterior cingulate cortex (PCC), using deterministic and probabilistic tractography. The PCC has been frequently called the central node of the default mode network (DMN), a resting state network that is believed to play an important role in maintenance of consciousness.
However, the two different tractography approaches yielded different results. For using ROIs with regions where fibers cross over may be useful to use a probabilistic tractography approach since it seems most effective in determining the underling connection. Though studies have shown that using the probabilistic approach may reveal reduction in structural connectivity paired nodes. This may be a consequence of using probabilistic tractography, which showed that with increasing tract length, the structural connectivity decreased, as reported by Khalsa et al. (2014). The results from Khalsa et al. (2014) indicated that the discernible decrease in structural connectivity in the PCC-mPFC could have resulted from the use of probabilistic tractography. Using different tractography approaches reveal different definitions of the functional connectivity strength and its relation with structural connectivity. An alternate to the probabilistic approach is the streamline. However both may complement each other with probabilistic tractography detecting more individual connections (especially where crossover is a problem) and streamline tractography demonstrating more consistent individual strength of connectivity” (Khalsa et al., 2014). Therefore, it might be beneficial to use multiple tractography approaches during analysis.

It would be interesting to see what factor may constrain the functional connectivity in the ROIs and how the relationship between the two connectivity measures changes depending on the type of analysis conducted.

More research should seek to quantify the dynamic changes that occur over time in a network structure during task-based fMRI or DTI studies. Further pursuit of mind. Research aimed in this direction might reveal that there are distinctive functional and structural connectivity profiles that do, indeed, have some predictive power of behavioral measures.
There might be a need to derive an alternate form of analysis that will be more encompassing and accurate for analyzing the correlational and possibly causal relationship between functional and structural connectivity between ROIs. Thus far, there does not seem to be an adequate model of analysis or algorithm that can capture all the various components involved in research endeavors like the ones previously discussed.
References


Flash-Lag Effect and Inverted Biological Motion

Chin-wei Andy Huang

The flash-lag effect (FLE) is an optical phenomenon that has been studied extensively to understand the underlying mechanisms which cause people to perceive flashed stimuli to lag behind a moving stimuli, even though the two appear at the same time. Although it is a well-studied phenomenon, it is still not fully understood what mechanisms underlie the effect.

Watanabe, Nijhawan, Khurana, and Shimojo studied the flash-lag effect in terms of how perceptual organization of the moving stimuli altered the effect size of the flash-lag effect (2001). In a series of eleven different experiments, Watanabe et al. observed that the perceived amount of lag differed relative to which edge of the moving object a dot had been flashed. Through different changes in shape and luminance, the researchers hypothesized that the flash-lag effect was heavily influenced by how the participants interpreted the perceived objects (as being one object or a series of separate objects). In many of the conducted experiments, Watanabe et al. found, with no significant difference between the results, that the leading edge of the object, regardless of the number of perceived objects, often are perceived to have more lag than the perceived lag in the non-leading edge. Thus, Watanabe et al. hypothesized that human vision, which includes the Flash Lag effect, “seems not to preserve accurate or precise information about spatial and temporal positions of stimuli... Rather, vision seems to preserve information about the field structure of images” (2001). Through their work, Watanabe et al. offers insight into the flash-lag effect as a process that may be strongly influenced by top-down processes.

In a previous study done the previous quarter, we found that, in terms of biological movement, there seemed to be a larger perceived flash-lag effect when participant saw a flashed
dot appear around the legs of a walker. From these results, we hypothesized that the underlying mechanisms for the flash-lag effect in the case of biological movement relied on the top-down influence of action prediction on the lower level processes engaged in visual perception. To test this hypothesis, we expanded on the work of Troje and Westhoff as they studied the effects of inversion on how participants perceive biological movement (2006). In their experiment, Troje and Westhoff not only found that the perception of feet act as one of the most important signifiers of biological movement, but also that the inversion of the feet into an unnatural position seems to result in less perceived visual accuracy. We extend this finding to the present study by hypothesizing that the perceived larger flash-lag effect found to affect people’s perception of the feet will not be as prominent when the walker is inverted.

Method

Participants

There were 5 participants in total, all of whom were students at the University of California, Los Angeles. Each of these students volunteered for the experiment using the SONA system. Upon completion of the experiment, each participant was given credit that could be used in his or her college courses.

Procedure and Design

The experiment was a two-way within-subjects alternative forced choice design. The two independent intervals were direction the walker was walking (either forward or backward) and the orientation of the walker (either right side up or inverted. Each participant was asked to complete a visual task where they were to judge the position of a dot relative to a moving object. In this case, the participants were asked to judge the perceived position of the dot relative to the
red limbs of a grey biological walker. They were told to hit the “left arrow key” when the dot appeared ahead of the moving limb and the “right arrow key” when the dot appeared behind of the moving limb. The participants could press the “down arrow key” if they had completely missed the stimuli. The experiment was also designed to have two sessions and they would only see one orientation of the walker during each of the sessions. The experimental trials was ordered in two blocks of 140 trials each and each block contained a walker that moved in a different direction. The order in which the participants experienced the orientation and the direction of the walker was counterbalanced to combat ordering effects.

The biological walker was made up of 8 different joint positions joined together to form the likeness of a walking figure. The walker’s right arm and legs were colored red and the dot would be flashed at temporally displaced offsets of the hand or the feet positions. There were 8 total positions that the green dot could appear ranging from a position of 66 ms (ahead of the limb) all the way to -132 ms (behind the limb), with every other point appearing at a 33 ms offset from each other between those two extreme points. A 0 ms differential between the red limb and the dot would mean that the dot was flashed on the position of the red hand or red feet.

At the beginning, each participant was asked to sit in front of the display monitor with their heads about 57 cm away, which was marked by a chin support. Once the participant was comfortable, the experimenter read the instruction of the experiment. After the participant indicated that he or she understood the task, the participant started the visual task, which started with a set of 10 practice trials. During the practice trial, the walker moved at a slower rate than the movement rate during the actual experimental trials. The experimenter remained in the room during the practice trial to answer any questions that the participant had and to check that the participants understood the task correctly. After the practice trial was completed, the participant
was instructed to move onto the actual experimental trials. Once both blocks of the experimental trials were completed for the first session, the participants were free to leave, but each was asked to come back the next week to complete the second session of the experiment.

The procedure of the second session was identical for the first session except that the participants were asked to complete a questionnaire to find the participant’s autistic quotient score. In an earlier study by van Boxtel and Lu (2013), researcher found that autism affects the way that people perceive biological motions.

**Results and Discussion**

The response rate averaged over the 5 participant for each body joint, body orientation, and walking direction is shown in Figure 1 below. Using separate analysis of variance tests on the different hand and foot joints from the different conditions yielded a significant two-way interaction effect for the foot position \( F(1,4) = 11.108, p = 0.029 \) and a significant main effect of walker walking direction for the hand position \( F(1,4) = 57.316, p = 0.002 \). The significant two-way interaction for the foot position acts as evidence for the proposed hypothesis that the inversion of a walker counteracts the perceived effectiveness of the flash-lag effect at the foot positions. In the right side up walker, participants tended to experience a more powerful flash-lag effect when the walker was walking forward than when the walker was walking background. However, this difference in flash-lag effect when the walker was inverted is not seen, which is evidence that the human visual system is sensitive to the proper orientation of the biological orientation, especially in terms of the position of the feet. If the feet of perceived biological motion is in a natural space, the action- prediction component of the flash-lag effect seems more pronounced. However, given an unnatural foot position (i.e., inverted feet), the flash-lag effect is
no longer subject to action-prediction as the movements of the feet become less easily predictable.

The significant main effect of the walker walking direction for the hand position demonstrates again that the ability to predict the subsequent position of a moving object seems to play a big part in the flash-lag effect. However, because these effects seem to only influence a certain joint position on the body, it would seem that the visual system might have different sensitivities to different body joints. This may be further evidence that there is increased attention paid to the leg joints of the body because they are more significant indicators of biological motion as suggested by Troje and Westhoff (2006).

For future studies, it may be interesting to study the influence of object perceptual organization on the flash-lag effect, especially in terms of a walking figure. It could be interesting to see the result when only one half of the walker is inverted, while the other remains upright and how that may influence the perceived flash-lag effect. Furthermore, such a study could further explore the importance of natural foot position on the flash-lag effect, which may lead to a better understanding of the link between action perception and the flash-lag effect.
Reference


**Figure 1.** Plots of the response rate that the dot appeared behind the moving limb when the walker was upright or inverted.
Nature Letter Assignment

Cody Sakurao

I. Introduction

Nearly everybody loves watching television, playing sports, or any other activity that requires your eyes in order to function. Vision is something that is constantly happening throughout the day, effortlessly firing neurons that tell your brain what you’re observing. But, how exactly does your visual system pick up your favorite celebrities every dance move? How do professional baseball players pick up the pitch within milliseconds in order to hit the ball?

All humans and other mammals have a centre-surround receptive field organization in their visual systems that assist in deciphering motion through space. This is made apparent in visual motion by the antagonistic interactions between centre-surround regions of the receptive fields in multiple direction-selective neurons in the visual cortex (Tadin et al. 2003). With this knowledge, Tadin et al. conducted series of psychophysical experiments in which they made the counterintuitive observation that “increasing the size of a high contrast moving pattern renders its direction of motion more difficult to perceive and reduces its effectiveness as an adaptation stimulus.” Tadin et al. also found that spatial antagonism of motion signals observed at high contrast turns into spatial summation as contrast decreases. These findings suggest that integration of motion signals over space depends on the visibility of those signals, resulting in the visual system correctly registering information.

II. Experiments/Conclusions
In the first experiment, threshold exposure duration required for human subjects to accurately identify the motion direction of a Gabor patch was measured. In the first condition, subjects viewed foveally presented Gabor patches of various sizes and contrasts. The spatial frequency and speed were held constant. Contrast of the Gabor patch was increased and decreased using a Gaussian envelope, which presented brief motion stimuli.

At low contrast, duration thresholds decreased with increasing size, which implies spatial summation of motion signals. This result was expected and consistent with earlier reports. Conversely, at high contrast, the results were so predictable. At high contrast, the duration thresholds increased as the Gabor patch width increased from 0.7° to 5°. For Gabor patches smaller than 2.7° performance improved with increasing contrast, while patches larger than 2.7° saw their performance substantially decline. This makes 2.7° a ‘critical size’, and also implies neural processes different than spatial summation. These surprising results led to several other experiments testing whether surround suppression was responsible for the worsened vision. Similar contrast-dependent size effects were obtained when the stimuli were faster and with Gabor patches whose bandwidth were held constant. In the second experiment, the contrast of the stimulus was manipulated by adding varying amounts of noise to a fixed-contrast Gabor patch. When motion appeared in high contrast noise, evidence for spatial summation was found. When Gabors appeared in weak noise or no noise at all, evidence for spatial suppression was found. Simply put, the presence of noise improved the visibility of large-motion stimuli. Effects of random dot displays presented in a spatial Gaussian envelope were also tested. Results showed that duration thresholds for these stimuli also showed evidence of spatial summation at low contrast and spatial suppression at high contrast, further bolstering their findings.
Since receptive fields of motion sensitive neurons increase with retinal eccentricity, Tadin et al. tested whether the detrimental effect of stimulus size at high contrast would diminish with increasing eccentricity. The display size was manipulated for a range of eccentricities and the contrast was held constant at 92%, but once again, foveal presentation results exhibited surround suppression. As eccentricity increased, duration thresholds decreased for all sizes of stimuli. This size dependent threshold relationship changed systematically with eccentricity, with no effect at the largest eccentricity tested.

The third experiment involved isoluminant motion in which there was a Gabor patch consisting of overlapping isoluminant red and green gratings. A phase shift technique was used by increasing phase shift from 0° to 90° of a fixed-contrast Gabor patch. Using this technique, thresholds were compared for isoluminant (red-green) and high luminance (yellow-black) Gabor patches. Luminance contrast stimuli were found to exhibit surround suppression, while threshold for isoluminant stimuli decreased with increasing size, exhibiting spatial summation. This is interesting because isoluminant motion with large stimuli was perceived more accurate than luminance-defined motion, most likely because the former isn’t affected by surround suppression.

A motion after effect (MAE) is an illusory perception of motion after a prolonged exposure to motion, which is thought to reflect the adaptation of motion-sensitive neurons. According to previous findings, adapting to a large high contrast moving stimulus should result in a weaker MAE. Tadin et al. tested this prediction by inducing MAE with moving Gabor patches of varying sizes and contrasts. As expected, MAE strength decreased with increasing size when subjects adapted to a high contrast Gabor patch, indicating spatial suppression.
Conversely, MAE strength increased when a low-contrast adapting stimulus of increasing size was used.

**III. Conclusions/Implications Critique**

This study showed that large objects destructively affect motion perception in humans. This finding goes against what is intuitive and questions research from previous motion perception studies. For instance, spatial summation is currently a basic characteristic of psychophysics yet this study showed that spatial summation is a characteristic of motion processing only in conditions of low visibility. Results also showed that dividing a large, high-contrast object into smaller parts improved performance in a speed discrimination task, with surround suppression as a possible explanation. Increasing the contrast of a briefly shown, large drifting gradient decreased performance. This may be because brief motion stimuli stimulate motion filters in opposing directions of the motion. Paired motion filters can’t explain the MAE results that generalize the findings to prolonged motion stimuli.

Tadin et al. came up with several reasons why their results could reflect the receptive field properties of centre surround neurons in the medial temporal visual area (MT). The first reason is that impaired visual performance with large stimuli is a characteristic of antagonistic centre-surround mechanisms. Secondly, the critical size at which strong surround suppression occurs can affect the surrounds of MT neurons with foveal receptive fields. This critical size rules out other cortical areas such as the primary visual cortex (V1) and part of the medial superior temporal (MST) due to their differing receptive field sizes. Third, the harmful effect of stimulus size disappears in the periphery, which is consistent with the increase in MT receptive
fields with eccentricity. Next, MAE is weaker if induced with large, high-contrast stimuli.
Lastly, MT neurons respond less to motion of isoluminant gratings than to motion of luminant
gratings. This correlates with the inability of isoluminant motion to produce surround
suppression.

IV. References

Tadin, Duje, Joseph S. Lappin, Lee A. Gilroy, and Randolph Blake. "Perceptual Consequences
Effects of Interactivity on Perceptual Learning in Mathematics

Emma Murray

With the advances in technology that we see today, there has been a dramatic change in the field of education and how we approach learning. One such change is the shift from passive learning as a teaching approach to active learning involving interactivity as a teaching approach. According to Michael Prince’s review of education literature, active learning is most commonly defined as a method of teaching in which the learner is actively engaged in the material (2004). This is counter to the more traditional passive methods of teaching in a lecture organization in which students simply experience the information as the teacher delivers it to them. In addition, there has been an increasing focus on perceptual learning as a way to increase learning, particularly through perceptual learning modules, and how the active and passive approaches can be applied to optimize the module for learning.

This shift from passive learning to active learning has come in the wake of a variety of studies demonstrating the benefits of active learning for long-term retention of the information. One such study showing this effect is that done by Roediger and Karpicke in 2006. In this study, they sought to extend the previous findings demonstrating the benefits of testing for long-term recall, or the testing effect, to the more specific domain of education. The study also controlled for the common concern that the results of the testing effect are simply due to extra exposure to the material by allowing the control condition to study the passage during the time that the experimental condition was being tested on the passage. Within the study there were two different experiments done to examine the testing effect in depth. In experiment one, participants were either in the control condition, in which they studied a passage that they would later be
tested on and were then allowed to study it again, or, they were in the experimental condition, in which they studied the same passage and then were tested on the passage in a free-recall style test without any feedback before receiving a final test. The results of a final free-recall test given with either a five minute, two day, or one week delay revealed that testing leads to increased long-term retention when compared to repeated studying. In other words, there was greater recall on the final test with a five-minute delay for those in the control condition than those in the experimental condition. On the other hand, the results were reversed for the final test with a two day delay, and even more so for the final test with a one week delay, such that recall was higher in the experimental group than in the control group (Roediger & Karpicke, 2006). These results confirm the advantage of active learning by demonstrating that active engagement with the material in the form of a test led to greater long-term retention, which is the ultimate goal of learning.

Experiment two sought to determine whether this testing effect was limited in its impact by examining if there was a difference in the effect of a single test versus multiple tests. To determine this, participants were in one of three conditions: the no-test condition in which they studied the passage four times, the single test condition in which they studied the passage three times and were tested once, or the repeated test condition in which they studied the passage once and were tested three times. The results revealed that the testing isn’t, in fact, limited because on a final test administered with a one week delay, subjects in the repeated test condition recalled significantly more than subjects in the single test condition, who in turn recalled more than subjects in the no-test condition. A similar pattern from experiment one was also replicated in experiment two in the difference between which conditions produced greater recall on the short-delay final test in contrast to which conditions led to greater recall on the long-delay final test.
On a final test following a five minute delay, participants in the no-test condition recalled the greatest amount of information from the passage, followed by participants in the single test condition who recalled more than those in the repeated test condition (Roediger & Karpicke, 2006). These results are particularly powerful because in an analysis of how many times the participants read the passage, it was revealed that those in the no-test condition read the passage many more times than those in the repeated test condition, but participants in the repeated test condition still recalled significantly more of the information (Roediger & Karpicke, 2006). This demonstrates the power of repeated testing, in addition to testing in general, to prevent forgetting, which is of great concern in the education field. These results provide support for the term “desirable difficulties” and the benefit of difficult processing to creating long-lasting memories for information. The authors of this study suggest that the reason that testing leads to increased recall is due to the matching of the testing condition during the study portion of the experiment with the final test condition. This match allows you to practice retrieving the information in the same way that you will need to on the final test, thus strengthening these retrieval pathways for later ease of recovery (Roediger & Karpicke, 2006).

The increases in technology mentioned earlier have led to a greater interest in how different learning approach theories, such as the active versus passive learning theory studied above, might apply to multimedia platforms. A study done by Evans and Gibbons in 2007 looked at this by examining whether what they refer to as the interactivity effect, based on the active-learning and passive-learning hypotheses, has an impact on learning from a computer-based module. Evans and Gibbons define interactivity as learning activities in which the learner is not simply receiving the information, but rather is participating with the information or the teacher of
the information (2007). In the context of this study, the learner is participating in the learning by interacting with the computer-based module.

The two hypotheses that can inform what effect interactivity with a computer-based module can have on learning are the active-learning and passive-learning hypotheses. The active-learning hypothesis suggests that interactivity will increase learning because the learner is more engaged with the information (Evans & Gibbons, 2007). On the other hand, the passive-learning hypothesis proposes that interactivity will have no effect on learning compared to non-interactivity, if the learner is receiving the same information in both settings (Evans & Gibbons, 2007). Similarly to the active-learning hypothesis, the study done by Roediger and Karpicke, described above, suggests that this interactivity would increase learning (2006).

To test these hypotheses, participants, with little previous knowledge of the topic, learned about how a bike pump works. They were divided into two conditions, interactive and non-interactive. In the interactive condition, there were three aspects of the module that allowed participants to be involved in the learning: pacing control, interactive self-assessment questions, and an interactive simulation. Pacing control referred to the subjects’ ability to press a button to indicate that they were done with the learning of one portion of the material and ready to move on to the next. Interactive self-assessment questions were questions that allowed the user to choose an answer from a variety of options, and actively drag that option to the answer box where they would receive feedback, or an opportunity to choose a different answer. Lastly, interactive simulation referred to an animation that allowed the learner to “operate” a bike pump to pump up a balloon by clicking to operate the pump (Evans & Gibbons, 2007). After the module, participants were tested on their knowledge using a test that required them to supply an answer to two questions that checked retention of the information, and three questions that
checked their understanding by looking at their ability to apply the information (Evans & Gibbons, 2007).

Using this procedure, they found that the participants who experienced the material in an interactive way scored significantly higher on the transfer portion of the test than those who experienced the material in a non-interactive way (Evans & Gibbons, 2007). This indicates that those participants who experienced the material interactively developed a deeper understanding of the material that allowed them to apply their learning to a different kind of problem at test. An additional finding showed that those participants who were in the interactive condition of the module took a significantly shorter amount of time to complete the questions on the final test (Evans & Gibbons, 2007). This may also indicate a deeper learning that comes from the interactive learning strategy, because a deeper understanding of information often leads to an ability to answer questions about that information more quickly. These findings are highly relevant in the increasingly technological culture that we live in because they illuminate the various ways that we can maximize the positive effects of technology on learning. In addition, these findings confirm the findings from Roediger and Karpicke, and solidify the idea of active, interactive learning as a way to increase retention and deep understanding, even through the platform of technology modules.

Given this basic understanding of the general difference in learning between active and passive teaching methods, the present study sought to apply this difference to the teaching structure of perceptual learning in the form of a perceptual learning module. Kellman, Massey, & Son (2009) define perceptual learning as improvements in information extraction as a result of practice. These “improvements in information extraction” can be broken down into two components that serve as advantages gained through perceptual learning: discovery and fluency.
Discovery refers to the ability to find features that are relevant to learning a specific pattern and classification. On the other hand, fluency refers to the ability to extract information that leads to making these discoveries, as described above, more quickly and automatically with practice and exposure to a large number of examples that vary systematically.

In looking at previous research involving the more specific cognitive cause of perceptual learning, there is some competing research that makes it difficult to determine conclusively how perceptual learning occurs. Kellman et al. (2009) suggests that perceptual learning occurs through a selection process, such that over many trials in which many varied examples are presented, invariant patterns are learned when some inputs that enter lower-level processing are selected as important and weighed more heavily so that they become information that enters higher level processing. These selected inputs are chosen out of the large number of total inputs that enter lower level processing and they are selected because of their many invariant occurrences in examples, despite the highly variant examples presented. This explanation for how perceptual learning occurs is presented in opposition to the idea that perceptual learning physically alters sensory detectors in order to help them to detect the invariant patterns in varying examples; however, the evidence associated with this description is much less conclusive (Kellman et al., 2009).

Because the present study covered perceptual learning in the domain of mathematics, it is important to discuss how perceptual learning can play a factor specifically in this domain. Because mathematics is a higher-level concept domain, the pattern recognition gained through perceptual learning is incorporated with declarative and procedural knowledge about the topic in order to gain a full understanding of the concept (Kellman et al., 2009). Kellman et al. explored this domain as well, and similar to the present study, used multiple representational formats to
express each math example that participants were scheduled to learn. They described the purpose of using multiple representational formats in a perceptual learning study as two-fold. First, participants would learn to perceive patterns, through discovery, and do so efficiently, through fluency, in each of the individual representational formats. Second, participants would learn to relate and perceive certain structural aspects across the representational formats and understand how a specific pattern in one representational format would be represented in another representational format (Kellman et al., 2009).

The present study was designed with the intention of determining whether different trial formats (active, passive, or passive-active combination) have an effect on perceptual learning in the field of mathematics. To achieve this, we gave participants a pretest covering the transformations of graphs, trained them on the subject using a perceptual learning module with one of the three different trial formats listed above, gave them a posttest to measure learning as a result of the training, and then finally, gave them a delayed posttest a week later to measure learning retention as a result of the training.

In looking at the predictions and results for this study, there were three different measures of interest: transfer accuracy, response time, and efficiency. Transfer accuracy looked at what accuracy was at test when compared with pretest, posttest, and delayed posttest. This measure essentially intended to look at the discovery aspect of perceptual learning. For this measure, we expected to find that accuracy would remain constant across the three conditions of active, passive, and passive-active. This was predicted because those participants in the active and passive-active condition completed the module only when they had learned a set amount of information and those participants in the passive condition were matched in how many trials they saw with a participant who had completed the active condition to the set amount of learning.
Because of this, all conditions saw the same material and presumably had learned all of it completely, thus they should all have been able to discover the patterns through perceptual learning.

Response time looked at what the overall average response time was for each correctly answered question. This measure essentially intended to look at the fluency aspect of perceptual learning. For this measure, we expected to find that fluency increased, so response time decreased, in the active condition because participants were practicing answering the questions that they would later be testing on; this allows for ease and speed of answering. This prediction is supported by the findings of Roediger & Karpicke (2006) about the nature of active learning in general and its ability to increase learning when compared with passive learning, whether in the context of perceptual learning or not.

Finally, efficiency looked at posttest gains (posttest – pretest) or delayed posttest gains (delayed posttest – pretest) divided by the total number of trials completed in the module. For this measure, we were only interested in comparing the active condition and the passive-active combination condition because the passive condition had a set number of trials and so the number of trials had no bearing on their learning and improvement throughout the module. When looking at this measure, we expected to discover that the passive-active condition was the most efficient for three reasons. First, having some passive exposure before completing the active module decreases the number of random, incorrect answers that generally occur at the beginning of the active module when participants are first beginning to learn the patterns. Second, the early correct answers that come from being primed are morale boosting and help participants to get more questions correct for the rest of the module. Third, having consistent correct answers from early on in the module allows you to move through the module more quickly.
References


Virtual Reality, Context Dependency, and Learning

Jacob Yu Villa

As technology improves, so does the potential to achieve greater things. One of the most important aspects improving is the effectiveness of delivering information. Virtual environments have shown much prospect in becoming a tool for enhancing the acquisition of information (Moskaliuk et al, 2012). Advancements in computer graphics and realistic simulation have broadened the capabilities of virtual environments. These capabilities range from training for occupations such as military or police positions to learning words of another language (Bailenson et al, 2008). With virtual environments becoming more commonplace with the growing popularity of video games, and likewise with immersiveness with the growing popularity of 3D, it is not unreasonable to predict other practices with immersive virtual environments becoming more ordinary.

According to Blascovich et al (2002), a virtual environment is one that provides “synthetic sensory information that leads to perceptions of environments and their contents as if they were not synthetic”. This means that anything ranging from head-mounted displays that help visualize an artificial environment to wearing earphones that help localize sound are possible features of a virtual environment (Bailenson et al, 2008). Playing computer games with a joystick or controller is another example of this. Virtual environments considered as immersive would then be ones where some special equipment lets the player directly control an avatar using their own movements. An example of equipment that does this would be Microsoft’s Kinect. Considering the lack of effective technology in this field, we are still a ways away from properly implementing true immersive virtual environments. Therefore, maximizing specific sensory
information would be currently the optimal method for immersion. This is what Joey and Nicco’s studies aim to do using the Oculus Rift.

One way that virtual environments are effective at delivering information is due to the ease of creating environmental context. Environmental context has been found to have positive effects on recall (Dalton, 1993). Being tested in the same context that one learned the material in improves the number of recalled items. Therefore, if one were able to easily create such environmental contexts, there could be powerful possibilities in regards to training and teaching. What makes virtual environments even more useful is the flexibility of design. Numerous virtual environments with great variety can be created more conveniently than if new environments (contexts) were created in the nonsynthetic world. Due to the massive potential and scope of virtual environments, one can create vast unfamiliar places that people can capture a more vivid first impression of. The unfamiliarity with the real world provides less room for interference of contexts, where they would clash and potentially lead to false memories. The inconvenience and amount of resources it would take to create the same, standout environments (such as fantasy or science fiction worlds) definitely make virtual environments a more feasible and attractive choice in context creation. The challenge then for virtual environment creators is to manufacture convincing worlds. It may still take time for technology to reach the point where virtual environments become completely indistinguishable from real environments. Meanwhile, increasing the sense of immersion through tools such as the Oculus Rift in OpenSimulator is enough to advance the research in this field at the current state of technology.

A technique that can be usefully integrated with virtual reality is the method of loci. This method improves recall of objects by imagining oneself at a familiar environment and thinking of a specific items to recall at certain landmarks. Then to recall one would just imagine walking
through the same environment using the same pathing thought up earlier. It becomes easier to memorize due to the associations made in this pathing. A virtual environment can bring this imagination to life by actually simulating the walking and placing objects to be recalled at landmarks. According to Moe et al (2005), the loci method needs three steps to work well: “a well-memorized loci pathway, the creation of good images of the items to be memorized and their adequate insertion in each locus”. Virtual environments can fulfill all these steps and increase the efficacy of this method by a significant amount due to being able to create more solid connections and encoding more effectively.

There are, however, certain limitations to the effectiveness of virtual environments. The reliance of studies using these environments on context-dependent learning exposes them to the same issues that context-dependent learning faces. While context may improve recall of items, it may not improve the acquisition of abstract knowledge (Smith, 1988). Furthermore, in order to better internalize the information gained, information should be decontextualized, so that what is essentially episodic memory is converted into semantic memory. It is not useful to simply remember facts; the mind needs to make connections in order to apply the information learned. This flexibility of thinking is one of humanity’s most powerful tools, so neglecting it would not prove favorable to the advancement of society. If virtual environments perfected our ability to memorize, it would still not have progressed much in the way of critical thinking. Better memorization and recollection of facts may be groundbreaking in terms of test-taking, but in the grand scheme of things needs to coordinate with the mind that makes the connections, that “decontextualizes” the information from episodic memory to semantic memory (Smith, 1988). This technique may cover some of the mind’s weaknesses, but without amplifying its strengths it simply is not as potentially powerful.
The applications for the usage of virtual reality for learning are large in number. Considering how much more and more humans are relying on technology, it is only logical that technology be used in order to maximize all aspects of life, including learning. The usage of virtual environments in order to enhance memory recall is a very promising method in theory, taking into account the current findings on the effectiveness of context dependency. Once enough studies have been performed to reliably conclude that virtual environments are effective at instructing, there can be major breakthroughs in the world’s training and education systems. As of yet there is still much to learn regarding both context dependency and virtual reality, but as studies advance we can grow hopeful in making an immensely improved future attainable.

My Experience this Quarter

I enjoyed working at the ALIVE lab this quarter. Granted, the current studies have yet to begin data collection, but preparing for them has been fun. I performed a test run as a participant for Nicco’s coin collection task, in which I collected coins in a virtual environment in a timed manner. I performed this task in front of a computer screen first and then with the Oculus Rift, which was still difficult at the time due to technical complications. I gave constructive feedback regarding the potential issues of motion sickness and reasons why. Despite the discomfort from feeling motion sickness, it was greatly enjoyable to test out the system.

Joey asked me to place arrows in her virtual environment to instruct future participants on the path they are supposed to take as they traverse throughout a science fiction world. In performing this task I was able to more clearly understand the effectiveness of the method, with the pathing specifically designed to indirectly bring attention to eye catching objects in the world.
Lastly, I determined the eligibility of participants by reading their surveys and matching them off a guideline of eligibility. Though reading the questions on the survey I could understand the basics of the requirements of the study as well as isolate control variables.

It was a lot of fun this quarter. Attending lab meetings was inconvenient due to a scheduling conflict with my classes, but the lab meetings during the previous summer were very interesting. Next quarter I hope to attend most if not all the lab meetings if there is no scheduling conflict, and I am utterly thrilled to see the current studies take flight so that I may become more involved with the research through assisting in data collection.

References


I began this quarter not knowing exactly what to expect out of my Psych 196B Research Project. All I knew was the topic, adaptive learning techniques in perceptual and fact learning for chemistry modules, which sounded very interesting. These adaptive learning techniques are based on the spacing effect, which is the increased learning that occurs when item presentations are spaced apart in time or order as opposed to crammed together much like the way many college students study (Sisti et al., 2007). Attempting to maximize the efficacy of this effect, the adaptive learning sequences use evaluative measures of accuracy and response speed to determine how well the subject knows the current item. Based on this information, the adaptive learning sequence then attempts to order the item presentations such that the items are spaced as far as apart as possible without being forgotten. The key idea here is that the more one has to work to remember the item, if successfully recalled, the greater increase in learning tensile strength and memory for that item. This concept is known as the “retrieval effort hypothesis” (Pyc & Rawson, 2009).

The official term for the adaptive learning schedule is the ARTS System (the Adaptive Response-Time-based Sequencing System). This system uses a priority score system, in which the priority for an item to reappear is computed as a function of accuracy, response time, and trials since the last presentation (Mettler & Kellman, 2014). The system also implements mastery criteria based on the same evaluative measures of speed and accuracy. Based on performance, the system automatically increases delay intervals for all items as learning strength increases. Since each item has a priority score which competes with the other items’ priority scores for presentation order, the system concurrently implements adaptive spacing for all learning items.
Though I did not initially have a good idea of what I would be investigating as a Research Assistant in the Human Perception Laboratory, my graduate student advisor, Everett Mettler, and I quickly came up with the concept of playing around with some recent data and attempting to analyze specific effects. Though this was my first foray into data analysis as an undergraduate student of UCLA, I became very excited at this prospect. The study that I started analyzing investigated the differences between adaptive and fixed learning schedules. There were 62 subjects from Collin College in McKinney, Texas. The study involved evaluating the differences between adaptive and fixed schedules in a fact learning task on chemistry modules. There were two conditions in the study: adaptive sequencing using the ARTS system, and a fixed continuously expanding schedule, which involves continuously expanding the delay intervals for each item at a fixed rate (i.e. 1, 5, 9,…). The experiment was split into four phases: 1) a pre-test phase of 24 questions that evaluated baseline knowledge of chemistry modules, 2) a training phase in which the subject was placed into either adaptive or fixed continuously expanding conditions and learned the chemistry modules to the best of their abilities for approximately an hour and a half, 3) a post-test phase consisting of 24 questions that measured the amount of learning that the subject underwent during the training phase, and 4) a delayed post-test phase one week (seven days) later to evaluate delayed retention of knowledge from the training phase. The stimuli presented were either ionic compounds or acids, and the subjects were presented either the name or the formula and asked to respond with the opposite in a multiple-choice task. Both directions (name to formula/formula to name) were presented within each subject.

As a beginner in data analysis, I did not know how to use the data analysis program that the lab typically used, so I did what I knew how to do and used MATLAB to rearrange the data and reorganize the trials into trials which were categorized as “Low Training Trials” (below the
average number of training trials for all category items of 8.1169) or “High Training Trials” (above the average number training trials of 8.1169). From here, I evaluated the post-test and delayed post-test accuracies for the category items labelled as low training trials vs high training trials. The resulting graph is pictured in Figure 1. As evidenced in Figure 1, there were some preliminary differences between the low training trials and high training trials categories on post-test and particularly on delayed post-test accuracies. This was an exciting preliminary finding, which gave me great hope for a more detailed analysis.

When I returned to Everett with my findings, we both got rather giddy and decided that further research needed to be done. However, he greatly assisted me in informing me that I should be analyzing individuals as a whole and categorizing individuals as “Low Training Trial Subjects” and “High Training Trial Subjects” rather than evaluating each individual trial. He also reminded me of the possibility of a priori knowledge of chemistry topics, which is evaluated in the pre-test phase of the experiment. Everett showed me that instead of looking at the post-test and delayed-post test accuracies I should be looking at the change scores (differences between
pre-test and post-test/pre-test and delayed post-test accuracies). Once I reformulated the data to account for all of these changes, I produced another graph, pictured in Figure 2.

Interpreting the data in Figure 2, one can see that the differences in post-test and delayed post-test accuracy remain between low training trials subjects and high training trials subjects. However, when adjusting for a priori knowledge (pre-test percent), we see that those who finished in less training trials tended to have more prior knowledge allowing them to retire all the items more quickly. Thus, when we look at the change scores, we see that the change percentages (differences between pre-test and post-test percentages) are almost identical. Interestingly, we do notice a small difference between low training trial subjects and high training trial subjects on change delayed percentages (differences between pre-test and delayed post-test scores), such that the mean of the low training trial subjects change delayed percent is 27.053% and for high training trial subjects it is 21.237%. Unfortunately, this finding was not significant, with a p-value of .175.
Upon returning to Everett with my results, he recommended that I chop up the data into four categories: Low Training Trials Subjects in the Fixed Condition, High Training Trials Subjects in the Fixed Condition, Low Training Trials Subjects in the Adaptive Condition, and High Training Trials Subjects in the Adaptive Condition. From this data split, we thought that we might be able to gain some insight into why some subjects were finishing more quickly than others and how they were performing on the change scores by condition. This data can be viewed in Figure 3.

**Figure 3:** Low training trials subjects vs. high training trials subjects separated by condition (adaptive vs. fixed) and evaluated in the following categories: pre-test percent, post-test percent, delayed post-test percent, change percent (difference between pre-test percent and post-test percent), and change delayed percent (difference between pre-test percent and delayed post-test percent).

Aside from the same trend of lower training trials leading to high change delayed percentages, there was one other interesting result. As is consistent with the main findings of the study, the subjects in the fixed condition tended to have slightly higher change scores on average than the subjects in the adaptive condition. This result was more pronounced in the high training trials...
subjects than in the low training trials subjects. What we can interpret from this (not significantly) is that when subjects require a large number of trials, a fixed schedule tends to be slightly more effective than an adaptive schedule at imprinting information in a memorable way such that change scores are greater on immediate and delayed post-tests are higher.

In order to look more closely at the trend mentioned above, I performed an additional analysis, pictured in Figure 4, looking exclusively at total number of training trials required for each of the four categories. In this analysis, I found very few differences in the total number of training trials between the low training trials subjects conditions: fixed -- 191.8 trials, adaptive – 196.9 trials. However, there was a significant difference in total training trials between the high training trials subjects conditions fixed – 329.4 trials, adaptive – 277.6 trials, p-value = 0.016.

![Figure 4: Total training trials in each of the four conditions: low training trials subjects for fixed condition, high training trials subjects for fixed condition, low training trials subjects for adaptive condition, and high training trials subjects for adaptive condition](image)

We can interpret this data in 2 ways: first, adaptive scheduling increases the speed of learning (number of trials until retirement) only for slow learners (subjects in the high training trials category). The second way we can interpret this data is related to the previous finding regarding higher change score performance of high training trials subjects in the fixed condition as opposed
to the adaptive condition. Since the high training trials subjects in the fixed condition experienced significantly more trials than those in the adaptive condition, it may be that those extra training trials in fact left behind a stronger imprint of the information, causing them to perform better on immediate and delayed post-tests.

In conclusion, my quarter of research in the Human Perception Lab was a beneficial experience. I got the opportunity to perform some extensive data analysis and programming. In my research I found several trends which are interesting. It appears that low training trials subjects tend to have higher change delayed percentages than high training trials subjects. Additionally, high training trials subjects in the fixed condition tended to have slightly (not significantly) higher change delayed percentages than high training trials subjects in the adaptive condition. However, high training trials subjects in the fixed condition also required significantly more training trials than high training trials subjects in the adaptive condition.
References


Pyc A, Rawson K. (2009). *Testing the retrieval hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory?* Journal of Memory and Language 60:437-447

The Study on Self-regulated Learning Through Learning Painting Styles

Jingqi Yu

Previous research has been devoted to the study of learning and memory. Earlier research overwhelmingly focused on the study of initial acquisition. However, as the definition of “learning” has broadened, recent researchers have begun to assess performance in long-term retention, transfer, and induction after the initial acquisition. One widely adopted approach is to study the effects of the representation of concepts on long-term learning (i.e., an interleaved schedule vs. a blocked schedule). Previous research has demonstrated the differences in performances between stages and maintained that different schedules could lead to distinct, or even opposite, results, depending on the various stages of learning (e.g., Schmidt and Bjork, 1992). Prior studies have suggested the advantages of interleaving in these later stages despite the fact that participants usually fared worse during the trials. A variety of studies on skill acquisitions (e.g., motor skills, language acquisition, and painting learning) provided evidence for this claim. For example, Shea and Morgan (1979)’s assessment of motor skill learning demonstrated that students who learned tasks under a random (interleaved) sequence of presentation performed better during both retention and transfer than those who learned under a massed (blocked) sequence, especially in transfer to tasks with greater complexity. After studying subjects’ performance in painting style learning, Kornell and Bjork (2008) found that, contrary to what was widely believed, even though learning paintings in a blocked order created a sense of fluency, spacing benefited induction more. These surprising findings were replicated. The underlying mechanism of this counterintuitive phenomenon has been examined. Researchers
found that the advantage of spacing in inductive learning is a result of highlighted differences between categories, not a mere result of temporal spacing (e.g., Kang and Pashler, 2012; Birnbaum, Kornell, Bjork, & Bjork, 2013). In addition, as Kornell and Bjork (2008) pointed out, because it seems easier to pick out the defining characteristics of an artist if one studies all of his/her paintings consecutively, learners tend to rate learning under an interleaved schedule as more effective. This natural tendency gives rise to the study on self-regulatory learning. Tauber, Dunlosky, Rawson, Wahlheim, and Jacoby (2013) concluded from their study on subjects’ learning of bird families that, even though numerous findings supported the benefit of interleaving for learning, an overwhelming preference for blocking was revealed. According to them, this could be due to several reasons, including participants’ tendencies to identify similarities rather than dissimilarities between categories. However, because little research has been done to investigate the effects of self-regulatory processes on learning, no conclusive statements can be made yet (use “can”! This is true, not hypothetical, and present tense). Thus, in addition to the study on the presentation of complex concepts, self-regulatory learning is a field worth studying.

The present study is interested in the participants’ self-regulated learning. It adopted the use of two schedules (interleaved vs. blocked). Different from previous style-learning studies that presented one of the conditions to the participants, participants were exposed to both schedules during the experiment. In addition, they were explicitly informed of the benefits of interleaving after finishing the first phase of the experiment: Interleaving enhanced the differences between the artists’ styles, and it is this discriminative processing that is crucial for the learning of different styles (Kang and Pashler, 2012). Considering this common bias, the experimenter also reminds subjects of this tendency and the reasoning behind it. As a result, the present study will
collect data conducive to the examinations of both the presentation of concepts and self-regulated learning.

**Method**

**Participants**

The experiment was launched during the middle of the quarter. During the time I was running the experiment, a total of 48 subjects who were enrolled in any psychology course at the University of California, Los Angeles participated in this experiment for course credits. The participants’ ages ranged from 18 to 22 years. Each subject served in one condition.

**Design**

The study was conducted as a between-subjects design. The independent variable was the method participants chose to study a set of new paintings during the second phase. Thus, the number of conditions was not fixed, as participants had the freedom to design their own learning schedule. The dependent variable was the correct number of artists recognized (out of 16).

**Material and Apparatus**

The experiment contained two phases, each containing a learning and a testing part. Each learning part composed of 48 pictures (eight artists × six paintings of each) and each testing part composed of 16 pictures (eight artists from the previous learning part × 2 new paintings of each). The eight artists selected during the first phase are different from the ones in the second phase. Therefore, a total of 128 paintings were incorporated into the design. Paintings were from artists such as Lewis, Juras, Mylrea and Petro. The selected paintings shared both similarities and differences. They had the same theme, landscape, so subjects could not rely on the theme to
make their judgments (e.g., when they saw paintings of ballerinas, they knew they were from Edgar Degas because ballet subjects were what he was famous for). Meanwhile, there were differences among each artist, including the use of brushes and the choice of paint (Figure 1). By having paintings under the same theme but differing in details, the task set a “desirable difficulty” (Bjork, 1994) level for the participants. During the first phase, the computer program decided the display order of these paintings. The practice schedule is neither a pure interleaving nor a pure blocked order. Rather, during each session, the paintings by each of the four artists were presented consecutively (in a blocked order) while the artistic works from the rest of the four artists were all mixed up together with those of other artists (in an interleaved fashion).

During the second phase, participants decided their own learning schedule. In between the two phases, a summary of the first phase would be provided. Specifically, it pointed out the differences between the two learning schedules (blocked vs. interleaving) and the benefits of interleaving to induction.

<table>
<thead>
<tr>
<th>Grossman</th>
<th>Hawkins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mylea</td>
<td>Margulis</td>
</tr>
</tbody>
</table>

[Images of paintings]
Figure 1. Examples of the paintings used during the experiment.

Procedure

Each participant was assigned into an individual room and was informed that the experiment was going to take 15 to 30 minutes. In the experiment room, participants were asked to follow the instructions given online. During the actual experiment, the instructions informed participants that they would be shown six paintings by each of eight different artists and their task was to learn each artist’s painting style. Throughout the learning experience, participants were expected to figure out artistic techniques unique to each artist to help them make decisions on the final test. In addition, after the learning part, they would be shown new paintings by the studied artists and asked to select, from a list of names, which artist painted each painting. Each painting was displayed for 3 seconds with the artist’s name shown below. Right after learning all
48 paintings, each participant had 45 seconds to play Tetris, followed by a test. During the test, participants were directed to the next page only after they made a decision, so they could spend as much time as they needed to recognize an artist’s painting style. However, because of this design, they did not have the chance to change their answers. After a debriefing, participants were directed to the second phase. In the second phase, participants were instructed to learn the six paintings of eight new artists with the method they considered more effective (i.e., the order they think would lead to the best learning of each artist’s style). To view a painting by an artist, participants were instructed to simply click on the artist’s name and each painting would last 3 seconds. Each painting was presented only once and the goal of learning was to associate a new painting with its painter on the final test. The button corresponding to the artist’s name disappeared after six clicks. Similar to the first phase, after learning 48 paintings, participants had 45 seconds to play Tetris, followed by the final test. The format of the final test was the same as that of the testing part in the first phase.

Results

We are still collecting data, as we need more participants. References


Creating desirable difficulties to enhance learning. In M. A. Gernsbacher and J. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society (2nd edition)*. New York: Worth.


The Influences that Guide Judgments of Learning

Joselyn Ho

Individuals with strong metamemory skills can accurately evaluate their memory capabilities and use that information to improve their learning and studying efficiency (Metcalf & Finn, 2008; Metcalfe, 2009). This ability is advantageous in a wide range of circumstances—for example, in the case of student learning, if students can be more aware of how well they can memorize and retrieve information, then they can employ better learning and studying strategies to retain more class material. Many researchers who conduct experiments about metamemory ask participants to provide judgments of learning (JOLs), which are numerical ratings that predict future memory performance (e.g., Rhodes & Castel, 2008). JOLs often do not align with actual memory performance, so comparing the JOLs to the participants’ recall results can give insight into the participants’ beliefs about their memory and the factors that influence those beliefs. The current study evaluates the effects of some of those factors, perceptual fluency and value-directed remembering, on JOL formation.

People’s tendency to underestimate extrinsic cues can lead to inaccurate JOL production (Castel, 2008). Extrinsic cues refer to the conditions of learning; examples of these include the number of words to be memorized and the serial positioning of each word, both of which are used in Castel (2008). Intrinsic cues, on the other hand, refer to an item’s attributes, such as its appearance, that can be thought to be responsible for the ease or difficulty of learning the item (Koriat, 1997). In each of the four experiments in Castel (2008), participants memorized a list of words, gave JOLs for each word, and underwent a free recall test. For the first experiment, as expected, participants’ recall of words followed the serial position effect described in Murdock (1962): recall of the first and last words of the list was significantly higher than recall of the
other words. However, their JOLs did not reflect this pattern. Thus, in this first experiment, the participants did not form their JOLs based on the extrinsic cues of word positions, possibly because they were unaware of them or did not pay attention to them. The participants must have placed greater priority on the intrinsic cues which led to the discrepancies between the JOLs and actual recall. JOLs aligned more closely to recall ability when the participants gave their JOLs prior to studying each word, were explicitly told the positioning of each word in the list, and underwent multiple study-test sessions. Thus, these participants, when given more obvious clues about the extrinsic characteristics, formed more accurate JOLs. By being forced to give JOLs before studying each word, the participants did not have any intrinsic cues to rely on; by being given access to serial position information, the participants could more readily utilize those extrinsic cues; and by going through multiple sessions, the participants could use their experiences to develop heuristics that synchronize their JOLs with their actual recall. The findings from Castel (2008), therefore, suggest that people usually do not pay attention to extrinsic cues during JOL formation unless under they are made more aware of those extrinsic cues. In most circumstances, then, intrinsic cues carry more weight in predictions of future memory performance.

JOLs that are made solely on the basis of intrinsic cues will often be incorrect because intrinsic characteristics can distract people from correctly understanding their memory capabilities. For example, item similarity can lead to encoding fluency-upwardly biased JOLs (Castel, McCabe, & Roediger, 2007). As expected, participants rated higher JOLs for identical word pairs such as “water-water” as better learned than non-identical word pairs such as “scalp-lunch,” and they also took less time to study the identical word pairs than the non-identical pairs. However, their recall for the identical word pairs was actually lower than for the non-identical
word pairs. These observations reinforce the effect of encoding fluency on the JOLs: reading the first word of each identical word pair expedites the reading of the second word due to the words’ perceptual similarity, which strengthens the perceived encoding fluency of those pairs. The participants overestimated the ease of recalling the identical word pairs because they believed that higher ease of processing leads to greater ease of retrieval. In actuality, the lack of associative strength between the identical words impedes memorization because the participants are forced to memorize which words among the list were not paired with an associative second word. This study’s findings support that people tend to use salient, intrinsic cues when forming their JOLs, despite whether those cues actually play a role in later recall.

Perceptual fluency, or how easily an item can be processed based on the item’s perceptual qualities, is another intrinsic cue that has been observed to influence JOLs (e.g., Rhodes & Castel, 2008). When people are shown words in large font sizes, they rate higher JOLs for those words. However, their recall ability does not depend on font sizes. The discrepancies between the JOLs and memory performance most likely result from the greater ease of processing the large words than the small words. Participants in Rhodes and Castel (2008) continued to rely on font sizes to form their JOLs even when they underwent multiple study-test sessions, were provided other intrinsic cues such as word relatedness, and were explicitly informed that font size has no effect on memory. Participants disregarded font sizes only when they studied words whose letters alternated between uppercase and lowercase letters. In this situation, the altered format increased the difficulty of quickly reading and processing the words, thus reducing the perceptual fluency of the words and the gap between the JOLs and actual memory performance. Based on these observations, perceptual fluency plays a significant role in
the prediction of memory ability even when it does not contribute to the ease of retrieval of stimuli.

People can use value-directed remembering to improve their metamemory when misleading intrinsic cues distract from accurate JOL formation. In value-directed remembering, people place greater significance on a stimulus and prioritize the memorization of that stimulus because doing so will be advantageous to their quality of life (Castel, McGillivray, & Friedman, 2012). A severe side effect of a medication could be one such stimulus, because the side effect would be valuable to remember if it is life-threatening. The current study combines value-directed remembering and perceptual fluency with the font size effect to investigate the relationship between JOLs and recall ability.

In the current experiment, participants view 20 side effects of a hypothetical drug and give a JOL for each side effect. The side effects are displayed in varying font sizes, and the participants are told that the size of the font indicates the level of severity of each side effect. After viewing the side effects, the participants undergo a free recall test. By assigning values to the words through the different font sizes, we can observe how the participants use those cues to guide their selection of which words to remember. We expect the JOLs for the side effects in large font sizes (more severe) to be higher than those for the side effects in small font sizes (less severe). Furthermore, if value-directed remembering occurs in this experiment, we also expect the recall of the large side effects to be greater than recall of the small side effects, due to participants placing more importance on the large, severe side effects. Exploring a possible effect of age could also be incorporated in the experiment by recruiting both young and old adults. Older adults most likely have more experience with medications and recognize the value of monitoring one’s health, so they may perform differently than the younger adults who have had
less exposure to drugs and do not yet understand the implications of severe side effects. We expect both age groups to have similar patterns in their recall ability but the older age group to have lower overall performance due to declining memory with age. We also expect the JOL ratings between the age groups to differ: For the severe words, the older age group will rate higher JOLs than will the younger age group. The JOLs, recall abilities, relationship between JOLs and recall, and potential age differences found in this study would have practical implications because doctors and pharmacists can use the information to improve their methods of informing patients about prescribed drugs or adjust treatment plans that involve side effects that patients perceive as severe. Knowledge of metamemory, therefore, applies to a wide range of practical circumstances, because memorization and retrieval of information occurs regularly and ubiquitously.

Metamemory research provides valuable findings that contribute to quality of life. Individuals who are aware of how their memory works can more strategically focus on important information. Due to JOLs’ sensitivity to intrinsic characteristics such as encoding and perceptual fluency, analyzing JOLs is one method of understanding people’s beliefs about their memory. One could argue that JOLs are not the most useful tools to examine metamemory because a laboratory setting differs greatly from real-life circumstances. However, JOLs still provide instrumental information about the factors involved in the judgment of memory capabilities, especially when an experiment incorporates complex concepts such as value-directed remembering. Future research can devise new techniques or build upon current ideas to continue the study of the relationships between people’s knowledge of their memory skills and actual recall ability.

References
Castel, A. D. (2008). Metacognition and learning about primacy and recency effects in free recall:


Intelligence has long been a subject of interest and controversy in psychology. From defining intelligence and its many subcomponents, to discovering what determines the vast differences in intelligence observed across the population, intelligence has been an ambiguous mystery. However, one thing has remained unambiguous and that is what some refer to the “puzzle of intelligence”. The puzzle describes the unanswered question: why do individuals who perform well at some things tend to also be good at numerous other things, and why do individuals who perform poorly at some things also tend to perform poorly at other things. In other words, what is the underlying quality or mechanism that allows consistency of performance across many different kinds of tasks. This paper aims to solve the puzzle of intelligence with the missing piece that was there all along, but perhaps fell underneath the table. The missing piece is perceptual learning, one of the main focuses in Professor Kellman’s Human Perception Laboratory.

One particular paper written by Williams et al. (2008) maps out the puzzle of intelligence and presents evidence contributing to the solution. Their data demonstrate the puzzle of intelligence by measuring performance on 14 different subtests of intelligence, ranging from vocabulary and arithmetic tests to picture completion and letter-number sequencing. The results show that almost all of the tests are positively correlated with performance on all other tests with the strength of correlation falling between .40-.70. This common variance is proposed evidence of overall general intelligence, which in this paper and
other papers within the literature is referred to as $g$. A $g$ score was calculated by averaging results from all of the subtests. Interestingly, $g$ scores generated by three different test batteries correlated .99, which serves as further evidence of the overall $g$ factor that underlies differences in intelligence across individuals. Looking for explanations and mechanisms for $g$, the researchers found that speed, or reaction time, was also positively correlated across subtests. That is, participants with faster reaction times on one test had faster reaction times on other subtests as well. Furthermore, the trend showed that the relative difference between an individual response time and the average response time increased as the difficulty of the task increased. This substantial and reliable pattern of processing speed stands as a likely and important factor in differences in intelligence, but does not explain all of it. If anything, the finding just makes the puzzle of intelligence more specific—why are some people consistently fast at subtests of intelligence and other consistently slow at subtests of intelligence, and why is their speed positively correlated with performance?

The paper also supplies some evidence of working memory being a major contributing factor to $g$. They argue that an individual’s ability to “organize and allocate attentional resources” and retain cognitive functioning in the presence of competing stimuli is an important aspect of $g$. However, they also conclude “working memory tasks are far form perfectly correlated with intelligence” (Williams et al., 2008). Therefore, although the paper presents two potential explanations for individual differences in $g$, these explanations are only small pieces to the puzzle. In order for the puzzle to be solved, we must change our perception of what the final piece might look like. The answer does not rely on static natural intelligent abilities, such as speed or working memory, but individual’s difference in learning ability. Although the paper suggests, “learning ability is a major
component of fluid intelligence”, the emphatic treatment of the suggestion is missing. Instead, they leave the question up to other researchers, to “determine which learning tasks predict individual differences in intelligence and which do not, and then identify the specific characteristics of these tasks that make such a prediction popular”. This paper will attempt to pick up where Williams et al. (2008) left off.

Perceptual learning encompasses the speed and working memory factors presented in Williams et al. and is presented here as a possible main, underlying explanation of differences in g. Perceptual learning is perfectly described in Kellman and Massey (2013) as the ability to “attune to the relevant features and structural relations that define important classifications, and over time, [to] come to extract these with increasing selectivity and fluency”. In more concrete words, perceptual learning can be broken down into two intricately related and reciprocal factors: discovery and fluency.

Discovery refers to the ability to identify and extract the important and relevant components of something. It also describes the ability to “weed-out” irrelevant and distracting stimuli, which is clearly analogous to working memory. Fluency, on the other hand, describes the ability to achieve this discovery faster and with less cognitive load, which is analogous to the processing speed factor discussed in Williams et al. (2008).

Furthermore, fluency and discovery reciprocate as perceptual learning abilities increase. For example, as fluency increases, less cognitive load is used, which frees up cognitive processes to engage in new discovery, which can lead to a higher perception and conceptualization of a given task. Fluency then increases as this new discovery becomes more habitual and effortless, and the cycle continues. As the cycle continues, expertise is developed. With evidence provided by the Williams et al. (2008) studies, this paper will present an argument for
perceptual learning as the underlying principle of general intelligence.

Williams et al. (2008) have already presented the impressive pattern between processing speed, task performance, and task difficulty. With processing speed already directly mapping onto the fluency component of perceptual learning, we will focus on the argument for discovery playing an equally important role in g. Firstly, the Williams et al. (2008) paper points out that individuals have correlational high performance on each subtest despite the vast differences in structure and content in each subset. This might be explained by the individual’s overall discovery ability. In other words, perhaps these individuals are proficient at looking at a problem or task and extracting the relevant clues or information necessary to solve the problem, with no specificity of the kind of task at hand. Additionally, the individual’s ability to enhance their discovery as the task continues might correlate with overall intelligence.

Williams et al. (2008) conducted an analysis in which the “g-loading”, or the amount that each subtest correlated to the overall g score, was calculated for each subtest. They found that the vocabulary test had the highest loading factor (.83). Although this finding does not seem immediately relevant to perceptual learning, it is important to consider the degree perceptual learning plays in language acquisition and understanding. Evidence of this can be seen in humans’ ability to read words even if they are scrambled. People perceptually “see” the word by extracting the important information, in this case the context or the first and last letter, and not by looking at the order of all the individual letters. Additionally, people can easily speak and write grammatically correct, but when terms and definitions such as adverbs and prepositions come into play, people commonly struggle. Therefore with language, we operate using high perceptual learning, with discovery effectively ignoring irrelevant features such as individual letters and grammatical rules, and it happening
so fluently we do not consciously realize that this is the case. The second highest subtest with relevance to \( g \) was a similarity test (.80). This test consisted of finding the ways in which two things are alike, which directly relates to the discovery of relevant, similar characteristics and thus perceptual learning. The better one’s perceptual learning ability, the better their ability to extract similar characteristics and ignore different characteristics. For instance, there are probably 1,000 things different between the two images, but only 5 things similar. A person’s ability to ignore these 1,000 differences and focus on the similarities will perform the task better, and faster.

Additionally, perceptual learning emerges very early on in development when infants begin categorizing the objects they see to newly learned words and concepts. For example, in order for an infant to classify different types of dogs they see into the category of “dog”, they need not to pay attention to the color or the fur or the size, but extract the relevant information, which is how they are similar—they bark, they have a tail, their behavior. It is these clues or similarities that determine category membership. Therefore, using perceptual learning to find similarities is something we do before we can even speak, and thus provides a possible explanation for why a similarity test is so highly correlated with general intelligence. The test with the least relevance to \( g \) is the digit span test. This test had individuals repeat a series of digits forwards and backwards. In this test, the individual has to pay attention to every digit, in other words, there is no sorting of irrelevant and relevant stimuli. Therefore there is not a lot of perceptual learning taking place and thus, according to the theory in this paper, explains why it has a low contributing factor to \( g \).

Additionally the finding that the difference between individual and average response times increased as difficulty of the task increased also presents an argument for the importance
of discovery. Perhaps a task was more difficult because it had more competing and irrelevant information, and thus, better discovery was needed. Therefore, it follows that as task difficulty increased, so did the demand for perceptual learning in order to effectively complete the task. Thus, good perceptual learners performed even better than average on these tests.

Of course, this argument for perceptual learning as the basis of differences in intelligence is mostly speculative, and the testing and measuring of perceptual learning remains to be conducted. To do this, performance on general perceptual learning tasks should follow the analyses outlined by Williams et al. (2008), such as correlating performance with performance on other subtests and the overall contribution to g. If the perceptual learning theory outlined in this paper is correct, perhaps the addition of the g-loading of processing speed and performance on perceptual learning tasks would equal the overall common variance of g. These perceptual learning tasks should follow the guidelines defined by Kellman and Massey (2013), 1) the task focuses on commonalities and variations in structure as its primary learning content. 2) consists of short classification trials with varied instances and feedback 3) learning comes from transactions with the input, and not verbal exchanges. Additionally, Williams et al. (2008) postulate, “to find that performance on two apparently diverse tasks is mediated at least in part by a single, more elementary, process is a step towards this long-range goal”. Perhaps more elementary processes of perceptual learning can be measured by eye tracking or mouse tracking. Do good perceptual learners look at all aspects of a problem equally? Do they jump back and forth a lot? Or do they focus on one area? All of these questions can be asked in order to measure differences in perceptual learning quantitatively.

The arguments presented in this paper suggest that individual differences in perceptual
learning are the basis for individual differences in intelligence. In other words, our overall intelligence is greatly determined by our perceptual intelligence. Perceptual intelligence seems to be a logical basis of intelligence because perceiving is the cognitive processes that humans have been doing since day 1 of life. The better we are at this basic process—the better it evolves from determining relevant features of objects in order to recognize and classify objects by name to determining relevant features of a problem or task—the more intelligent we are.

References


The Effect of Movement Control and Navigation Control on Spatial Knowledge Tasks

Kylie Springsteen

People say they learn the lay of the land better when they are the driver, as opposed to being the passenger. However, there are two factors at play in this type of situation-- movement control and navigation control. Movement control would be whether you are the one in control of the movement, that is, if you are the driver or the passenger. Navigation control is whether or not you are actually making navigation decisions. Navigation can be active or passive; an active navigator makes decisions in forging a path and a passive navigator follows the lead of someone else and is uninvolved in the decision making. Rul von Stulpnagel and Melanie C. Steffens explore which of these factors is actually affecting spatial learning and/or what combination of these factors best aids spatial memory in their paper “Can Active Navigation Be as Good As Driving? A Comparison of Spatial Memory in Drivers and Backseat Drivers” (2012).

Previous research has shown that control of decisions and active, self-directed, and free exploration enables superior spatial learning (Péruch, Vercher, & Gauthier, 1995). However, there are two main problems that von Stulpnagel and Steffens target in the present study. One is that most previous research has been done in a virtual environment, so they conducted their study in a real environment to see if previous findings hold in a real environment. If so, this gives greater validity to these results and, in addition, gives greater validity to all virtual reality studies. A second problem they target is related to the fact that there are three distinct types of spatial knowledge; landmark knowledge is knowledge about distinctive and stable features of the environment, route knowledge is knowledge of the order of appearance of landmarks and turns on a given route, and survey knowledge is the formation of a cognitive map including features of
space and orientation. When administering a spatial learning task, it is important to consider what type of learning is occurring and what type of learning you are testing. Depending on how you operationalize active manipulation, it could be confounded with your object of measurement. For example, you may have your participants actively navigate by using maps. Maps have been shown to emphasize survey learning so if you are using maps and giving a landmark test, you may miss significant results. Previous research fails to acknowledge these distinctions and von Stulpnagel and Steffens carefully accounted for these possible confounds in their study.

Their study consisted of three experiments in which they manipulated movement control and navigation control and measured spatial knowledge in a set of tasks that specifically targeted each of the three types of spatial knowledge (von Stulpnagel and Steffens, 2012). In Experiment 1, participants navigated an unfamiliar park via maps on tandem bikes. A map of the park was divided into six sections. Participants were given a section, instructed to memorize the shortest path from Point A to Point B, and then to bike the path. When they got to Point B, they were given the map for Section 2 and repeated the same procedure. This was repeated for all 6 map segments. They were then given a distracter task, followed by a series of spatial memory tasks.

To manipulate movement control, participants were randomly assigned either to be the driver or the passenger of the tandem bike. To manipulate navigation control, participants were randomly assigned either to an active or passive navigation group. Active navigators were given maps which only labeled Point A and Point B, which required them to decide what the shortest route was. Passive navigators were given maps with the shortest path marked directly on the map. Therefore, passive navigators were not required to make navigation decisions, but simply follow the instructions given. This design resulted in a comparison of 4 groups- active driver, passive driver, active passenger or passive passenger.
They hypothesized that active navigators (using maps) would be better on survey tasks regardless of movement control, but worse on landmark tasks. Previous research showed that map activity leads to better learning of survey features (Münzer, Zimmer, Schwalm, Baus, & Aslan, 2006). However, existing research on landmark knowledge is mixed so they postulated that landmark learning might be worse simply because participants would be inclined to encode survey features due to the use of maps and will not pay their attention to landmarks.

In Experiment 1, they used four tasks to measure spatial learning. The first was a Tour Integration task, which measured survey learning. In this task, participants were given a map of the whole park and asked to draw the entire route they had taken, which required them to orient the segments correctly. Second, landmark knowledge was measured using a Landmark Recognition task. In this task, participants were given a set of 12 landmarks (6 they had seen and 6 they had not seen) and asked to indicate whether or not they had encountered that landmark on their ride. Third, a Free Orientation task was implemented. In this task, participants were placed in a random location and asked to find their shortest way back to the starting point. Fourth, a Landmark Pointing task was used, which asked participants to estimate, from 0 to 180 degrees, what direction a specified landmark was located.

The results of Tour Integration showed a main effect of navigation, that is, active navigators had acquired superior survey knowledge than passive navigators. The Landmark Recognition also showed a main effect of navigation, in which active drivers showed significantly worse performance than active drivers. The Free Orientation and Landmark Pointing tasks were too easy and too difficult, respectively, and did not show any significant results. Overall, these results were consistent with their hypothesis.
From here, they hypothesized that perhaps movement control may be advantageous in a familiar environment. Experiment 2 used exactly the same procedure as Experiment 1, this time in a familiar environment. They expected the results of Experiment 1 to hold true, that is, they hypothesized that active navigators would do better on survey tasks and worse on landmark tasks. However, they also hypothesized a main effect of movement control, that is, drivers would show better performance because driving requires the participant to pay attention and be aware of their surroundings, so they will encode more information about their environment. Also because a familiar environment requires less resources for orientation and navigation, this frees up attentional resources to encode environmental features and because driving inherently requires attention to the environment, spatial learning will be increased.

They measured learning using the same Landmark Recognition, the same Tour Integration task, and an additional Landmark Ordering task. In the Landmark Ordering task, participants were given a collection of landmarks and asked to order them based on the order in which they encountered them.

The Landmark Recognition test showed a main effect of movement control, that is, drivers performed better than passengers and no main effect of navigation. Landmark Ordering showed no effects. Tour Integration also showed a main effect of movement control, that is, drivers performed better than passengers and there was also no main effect of navigation. Additionally, there was an interaction. These results indicated that movement control may be influential to spatial learning in a familiar environment. However, further analyses of the interaction showed that the only condition that was significantly lower was passive backseat drivers. The other three conditions did not significantly contrast each other, so doesn’t show that drivers possess an advantage because even backseat active navigators performed comparably
well to drivers. Therefore, von Stulpnagel and Steffens (2012) did not attribute these results to an effect of movement control.

Experiment 3 was a little different. Von Stulpnagel and Steffens (2012) thought that one potential problem with their setup may be that the active and passive groups were encoding information differently and maybe the difference in their results were due to different encoding strategies, not different navigation or decision-making. They hypothesized that if this was the case, then the active navigation group would have more complex representations than passive navigators and the complex representation is what enabled their performance, not active decision making. To test this, they presented participants with 6 map segments and told them to memorize the shortest route from Point A to Point B (same setup as Experiments 1 and 2), however, after each map segment, they were asked to draw a picture of the map rather than navigating the route. They defined a simple drawing as one that only included intersections of the shortest route, and a complex drawing as one that included parts of the road system not related to the shortest path as well as any other landmarks, etc. Results showed that the active group had more complex maps than the passive group.

Then they looked at the results of the Tour Integration task. In Experiments 1 and 2, where participants actually navigated the route, active navigators did better than passive navigators on the Tour Integration task. If it is true that this difference was due to a difference in encoding rather than the actual navigation, then the difference between groups on this task should remain in Experiment 3. If it is not a matter of encoding strategy but actually is an effect of navigation, then we should not see a difference here since the participants in Experiment 3 didn’t actually navigate. Results showed no difference between groups on the task, which supports the idea that active navigation is the factor affecting spatial learning.
Overall, this series of experiments confirms previous research that shows that active navigation is more central to spatial learning than movement control. It also shows that research in a virtual environment is indicative of real life phenomena, which can have greater implications for virtual reality experiments, in general. This can be really important because some experiments are very difficult to run in a real environment, so if it’s shown that virtual environments reliably report real-life results, it makes these experimental results more valid.

Further research can explore some additional items. First, they can tease apart the Tour Integration task. The Tour Integration task also includes route knowledge, because the participant is required to draw the route the cycled, and recall the order of their decisions. This experiment did not successfully determine how much of these results are indicating the acquisition of survey versus route knowledge, so this can be further explored. Another change that can be made in further research is the inclusion of a true passive group, that is, participants are never shown a map and are simply along for the ride. This may be more indicative of a real-life scenario and might more accurately represent the amount of information a passive rider picks up. Another idea for future research would be to further explore why we are seeing different results between a familiar and unfamiliar environment. Von Stulpnagel and Steffens (2012) postulate that it’s due to attentional resources, but this is a hypothesis that can be experimented further.
References


Effects of Interactivity, Sequencing and Retirement on Factual Learning

Lorena Hinckley

Historically, text has been the dominant format for teaching factual material and books have been the primary instruction tool. However, with the advancement of technology, computers now have the potential to improve students’ learning. The question now is how should these computer programs be designed? Unfortunately, many computer-based teaching programs today have been designed to essentially be high-tech textbooks, displaying large amounts of information solely in a textual format. This paper explores the benefits of presenting information in an interactive way that engages the learner and asks the question of whether variables such as adaptive sequencing, the retirement of items, and the amount of exposure to items play a role as well.

But what does it mean for a program to be an “interactive”? In the literature, the term interactivity has been used to describe a variety of interactions including those between the teaching material and the student, between the teacher and the student, and between the students themselves (Moore, 1989). In computer-based learning programs such as the ones discussed in this paper, interactivity primarily takes the form of initiation, then response, and then feedback (Evans & Sabry, 2002). According to this model, each of these steps of the computer interaction involves an exchange of information between the learner and the computer program. Thus, in this paper we define a learning program to be interactive if it uses this computer-initiated interactivity as an essential piece of the lesson.

Previous studies have suggested that student interaction is essential in learning, supporting the teaching-as-communicating view over the teaching-as-transmitting view. The
view of teaching-as-communicating fosters the idea that it is not the teacher’s job to simply transmit information as in the teaching-as-transmitting view, but instead to encourage the learner to construct their own mental representations (Mayer, 1992). Therefore, interactivity according to this view should enhance the level of learning as it engages the learner and forces them to actively process the information being presented. For example, Moreno, Mayer, Spires, and Lester (2001) examined the relationship of interactivity and learning of science facts. By varying the level of interactivity in their computer program, they found that students who were allowed to actively participate in the program retained more information and were more successful at applying the knowledge to difficult transfer problems. This suggests that an active-learning environment is beneficial to the learning process. In 2007, Evans and Gibbons examined whether the addition of interactivity to a computer-based learning program would enhance the level of learning. They found that adding interactivity to computer-based learning programs increases both the depth of learning and the ability of learners to transfer the knowledge they have acquired to other tasks, resulting in higher accuracy and speed on transfer tests.

In the experiment described below, we examine the accuracy of the active-learning hypothesis which predicts that learning in an interactive setting increases learning through engaging the learners and making them actively participate in the learning process. This active-learning hypothesis is based on constructivist models of learning that argue the learning process must involve the learner in constructing their knowledge on a particular subject on the basis of what they already know and the new information they are receiving (Mayer, 1999). As students learn, they play an active role in the processing of new information such that when they are required to interact with a learning environment through choosing when they learn (e.g. clicking a button) or selecting what information they receive (e.g. by selecting from a list of options),
they develop an active relationship with the material presented. Another basic difference between actively presented material and passively presented material is simply the level of attention the learners pay to the material. When a student is active in the material they are learning, they are participating in active listening and because of this tend to learn better (Coakley & Wolvin, 1997). As a result, the active-learning hypothesis predicts that learning should improve when students use interactive learning programs.

Also examined in the present study is the effect of adaptive sequencing and retirement on long-term recall of factual information and training efficiency. Mettler, Massey, and Kellman (2011) describe an adaptive sequencing system (Adaptive Response Time Based Sequencing—ARTS) that uses both accuracy and response time (RT) as direct inputs. ARTS utilizes response times to assess learning strength and determine mastery, thus making fluency and accuracy the goals of learning. ARTS uses the equation \( P_i = a(N_i - D)[b(1-\alpha_i)\log(RT_i/r) + \alpha_iW] \), which relies on the rapid response of missed items, interleaving/enforced delay, dynamic spacing based on response time, and retirement criteria to determine the sequence of items in the learning system. Rapid response of missed items ensures items answered incorrectly are presented again more quickly by assigning them higher priority weight. Interleaving/Enforced Delay prevents the presentation of an item while its answer is still in working memory. Dynamic Spacing Based on RT means that as the learner gives faster accurate responses, the system uses response time on these correct answers to increase spacing between presentations of that item; the more accurate responses for an item, the less often the learner sees it. Lastly, ARTS uses retirement criteria based on both accuracy and response time to remove items from the learning set once the learner has mastered them. According to Mettler and Kellman (2014) and Mettler et al. (2011), adaptive
sequencing increases the accuracy and efficiency of learning. This adaptive sequencing will be used in the present set of experiments.

**Experiment 1 - Interactivity**

In this experiment, participants study a set of 24 African countries. For one group of participants the countries are be presented in an active format, for another other group the study session is set up as a passive learning task, and a third group sees a combination of the two. All participants complete the same geography test before the learning phase and immediately after completing the learning phase, and then take a delayed post-test one week later.

**Method**

**Participants**

In the experiment, 60 undergraduate students (age range 18-34) were recruited from the University of California, Los Angeles. The initial 17 participants participated between 2007 and 2008, and an additional 3 participants were run in 2014. Participants received course credit as compensation.

**Apparatus and Stimuli**

The stimuli in this experiment are slides depicting a map of the African continent with one of 24 possible countries highlighted. The countries are selected to be relatively unfamiliar ones (e.g. Lesotho, but not Egypt). To the right of the map is either a list of the names of the 24 countries in alphabetical order or the name of highlighted country (see Figure 4).

**Design and procedure**

This experiment had three conditions manipulated through which module the subjects are given: Sequential Interactive Retire (SIR), Sequential Passive-Interactive Retire (SPIR), and Fixed Passive No-Retire (FPN). The effect of these different modules on the performance of
participants on the post tests was measured in terms of accuracy and response time, measuring recall ability and fluency of the material. All three modules consisted of a pre-test, a training phase, a post-test, and a delayed post-test.

The pre-test and the post-test were identical in format. They each consisted of 24 randomly ordered questions (one per country). For each question, a map of Africa with one country highlighted was displayed on the screen, and the participants were asked to select the correct name of the highlighted country from the list of countries on the right. Participants were not provided with feedback. The pre-test was given immediately before the training phase and the post-test was given immediately after. The delayed post-test consisted of 72 questions made up of the questions of the pre-test presented three times.

During the training phase, participants completed a learning module dependent on their condition. In the SIR group, participants were presented with items in an active format, meaning when a country is presented, participants were instructed to select the correct name of the highlighted country. Immediately after choosing a country name, the participant was given feedback. The stimuli were blocked into groups of ten, and were presented in an order determined by an algorithm that adapts based on the participants accuracy and response times (Mettler et al., 2011). Through the retirement, items could be taken out of the learning set either by reaching the learning criterion (correctly identifying each country three times in a row in under 10 seconds per item), or through “quick-retire” if the learner correctly identified it on its first presentation in under five seconds. If a retired country was selected as the answer for a non-retired country, it was un-retired and brought back into the learning set. Participants were trained until they reached learning criterion for each country.
In the FPN group, participants were presented stimuli in a passive format, meaning the correct name for the country highlighted in the list of countries and written above the country on the map. Stimuli were presented in a random order for five seconds before disappearing from the screen, after which participants were prompted to continue. In this group, items were not adaptively sequenced and there was no retirement. Participants in this condition were trained until they reached 192 trials, regardless of if learning criteria were met. In the SPIR group, participants saw each country first in a passive format as in the FPN group, then continued the training in the active format identical to the SIR group, stopping once they had reached learning criterion for each country.

Results
Accuracy

Figure 1 presents the average accuracy of participants’ responses on the pre-test, post-test, and delayed post-test. From the pattern of results shown in Figure 1, all conditions appear to do equally well on the pretest, but on the post-test and delayed post-test participants in the sequential order interactive retire condition appear to do better than participants in the sequential passive/active retire condition and fixed passive no-retire conditions. Fixed passive no-retire appears to have the lowest accuracy on both the post and delayed post-tests.

The data were analyzed using a 3 (conditions: Sequential Interactive Retire, Sequential Passive/Interactive Retire, Passive Fixed No-Retire) x 3 (phases: pre-test, post-test, delayed post-test) repeated measures analysis of variance (ANOVA) to compare the effect of interactivity on the accuracy of responses. Mauchly’s test indicated the assumption of sphericity had been violated $\chi^2(2) = 10.21, p = 0.006$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.86$). This revealed a significant main effect of
condition \( (F(2, 57) = 8.21, \ p = 0.001, \ \eta^2 = 0.22) \), a main effect of phase \( (F(1.71, 97.72) = 798.95, \ p < 0.001, \ \eta^2 = 0.93) \), and a significant interaction between phase and condition \( (F(3.43, 97.72) = 7.21, \ p < 0.001, \ \eta^2 = 0.20) \). At pre-test there was no difference in accuracy of responses between conditions, \( p > 0.05 \).

Across all conditions, participants improved significantly from pre-test \( (M = 0.06, \ SD = 0.04) \) to post-test \( (M = 0.80, \ SD = 0.18) \), paired-\( t \) \( (59) = -32.16, \ p < 0.001, \ d = -5.68 \).

Participants also improved significantly from pre-test to delayed post-test \( (M = 0.60, \ SD = 0.19) \), paired-\( t \) \( (59) = -22.40, \ p < 0.001, \ d = -3.93 \). There was a significant drop in accuracy from post-test to delayed post-test, paired-\( t \) \( (59) = 13.49, \ p < 0.001, \ d = 1.08 \).

On average across all phases, the accuracy of responses in the sequential interactive retire condition \( (M = 0.51, \ SD = .06) \) was significantly lower than in the sequential passive/interactive retire condition \( (M = 0.54, \ SD = 0.08) \), the accuracy of responses in the sequential interactive retire condition was significantly higher than in the passive fixed no-retire condition \( (M = 0.41, \ SD = 0.16) \), and the accuracy of responses in the sequential passive/interactive retire condition was significantly higher than in the passive fixed no-retire condition, \( (F(2, 57) = 8.21, \ p = 0.001, \ \eta^2 = 0.22) \).

At pre-test there was no difference in accuracy of responses between conditions, \( p > 0.05 \). There was a significant difference in the post-test scores for participants in the active format \( (M = 0.86, \ SD = 0.08) \) and participants who saw trials passively \( (M = 0.68, \ SD = 0.25) \), \( t \) \( (38) = 3.02, \ p = 0.005, \ d = 0.97 \). At delayed post-test, participants in the active format \( (M = 0.62, \ SD = 0.12) \) also scored significantly higher than participants who saw trials passively \( (M = 0.49, \ SD = 0.24) \), \( t \) \( (38) = 2.18, \ p = 0.04, \ d = 0.69 \). When comparing active with active/passive presentation, there was only a slightly significant difference at the delayed post-test for participants’ accuracy
in the active format compared to participants in the active/passive format (M = 0.70, SD = 0.14), t (38) = -2.02, p = 0.05, d = -0.61. The accuracy scores of participants in the passive and active/passive groups differed significantly at both post-test (t(38) = 3.20, p = 0.003, d = -1.04) and delayed post-test, (t (38) = 3.44, p = 0.001, d = -1.07). This suggests that the level of interactivity of the module affects the accuracy of responses such that pure active format performs more accurately than pure passive format, but a combination of the two is superior to either of them on their own.

**Response Time**

The data were analyzed using a 3 (conditions: Sequential Interactive Retire, Sequential Passive/Interactive Retire, Passive Fixed No-Retire) x 3 (phases: pre-test, post-test, delayed post-test) repeated measures ANOVA test to compare the effect of interactivity on response times of correct responses. Mauchly’s test indicated the assumption of sphericity had been violated $\chi^2(2) = 106.04, p < 0.001$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.54$). The results show no significant main effect of condition ($F(2, 55) = 0.09, p > 0.05, \eta^2 = 0.003$), no main effect of phase ($F(1.08, 59.15) = 2.31, p > 0.05, \eta^2 = 0.04$), nor a significant interaction between condition and phase ($F(2.48, 59.15) = 0.75, p > 0.05, \eta^2 = 0.03$).

**Experiment 2 - Sequencing**

Experiment 2 was designed to examine the effects of sequencing on long term retention of factual information and fluency of fact recognition. In experiment one, the some conditions were adaptively sequenced while others were not, so sequencing could have been a possible confound.

**Method**
Participants

In the experiment, 40 undergraduate students (age range 18-34) were recruited from the University of California, Los Angeles. The initial 17 participants participated between 2007 and 2008, and an additional 3 participants were run in 2014. Participants received course credit as compensation.

Apparatus and Stimuli

The stimuli in this experiment were identical to those in Experiment 1.

Design and Procedure

This experiment had two conditions manipulated through the use of a sequencing algorithm during the learning phase: Sequential Interactive Retire (SIR) and Random Interactive Retire (RIR). The effect of these different modules on the performance of participants on the post-tests was measured in terms of accuracy and response time, measuring recall ability and fluency of recall. Both of the conditions consisted of a pre-test, a training phase, a post-test, and a delayed post-test identical in format to those in the Experiment 1. During the training phase, participants completed a learning module dependent on their condition. Both conditions saw the countries presented in an active format without sequencing (for information on active format and sequencing, see Experiment 1). In the training phase SIR condition, items were presented in an order determined by the sequencing algorithm described in Experiment 1. Items were presented in a random order in the training phase of the RIN condition. In both conditions participants were trained until they reached certain learning criterion, the same as in the previous experiment.

Results

Accuracy
Figure 1 presents the average accuracy of participants’ responses on the pre-test, post-test, and delayed post-test. From the pattern of results shown in Figure 1, all conditions appear to do equally well on the pretest, but on the post-test and delayed post-test participants in the sequential order interactive retire condition appear to do worse than participants in the random order interactive condition.

The data were analyzed using a 2 (conditions: sequential order interactive retire, random order interactive retire) x 3 (phases: pre-test, post-test, delayed post-test) repeated measures ANOVA test to confirm the effect of sequencing on the accuracy of responses. Mauchly’s test indicated the assumption of sphericity had been violated $\chi^2(2) = 13.59, p = 0.001$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.77$). Analysis revealed only a significant main effect of phase ($F(1.53, 58.14) = 888.97, p < 0.001, \eta^2 = 0.96$) and a main effect of condition ($F(1, 38) = 4.26, p = 0.05, \eta^2 = 0.10$).

Participants improved significantly from pre-test (M = 0.06, SD = 0.04) to post-test (M = 0.88, SD = 0.09), paired-t(39) = -57.81, p < 0.001, d = -11.77. Participants improved significantly from pre-test to delayed post-test (M = 0.66, SD = 0.15), paired-t(39) = -24.25, p < 0.001, d = -5.47. There was a significant drop in accuracy from post-test to delayed post-test, paired-t(39) = 10.16, p < 0.001, d = 1.77. The accuracy of responses in the sequential interactive retire condition (M = 0.51, SD = .06) was significantly lower than in the random interactive retire condition (M = 0.56, SD = 0.08), ($F(1, 38) = 4.26, p = 0.05, \eta^2 = 0.10$). This suggests that adaptive sequencing of materials in the sequential order interactive retire did not improve accuracy of recall as compared to those in the random order interactive retire condition.

Response Time
Response times of correct responses were analyzed using a 2 (conditions: sequential order interactive retire, random order interactive retire) x 3 (phases: pre-test, post-test, delayed post-test) repeated measures ANOVA. Mauchly’s test indicated the assumption of sphericity had been violated $\chi^2(2) = 67.97, p < 0.001$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.54$). The results show no significant main effect of phase ($F(1.09, 41.29) = 3.11, p > 0.05, \eta^2 = 0.08$), no significant main effect of condition ($F(1, 38) = 0.02, p > 0.05, \eta^2 = 0.0$), nor a significant interaction between condition and phase ($F(1.09, 41.29) = 0.24, p > 0.05, \eta^2 = 0.01$). There was no difference in participants’ response time at pre-test, post-test, or delayed post-test between conditions, $p > 0.05$, suggesting that adaptive sequencing of materials in the sequential order interactive retire did not improve the response time on either post-test or delayed post-test as compared to those in the random order interactive retire condition.

**Experiment 3 - Retirement**

The purpose of Experiment 3 was to determine the potential effects of retirement during the learning phase. Retirement is defined as the removal of an item from the learning set.

**Method**

**Participants**

In the experiment, 40 undergraduate students (age range 18-34) were recruited from the University of California, Los Angeles. The initial 17 participants participated between 2007 and 2008, and an additional 3 participants were run in 2014. Participants received course credit as compensation.

**Apparatus and Stimuli**

The stimuli in this experiment were identical to that in Experiment 1.
Design and Procedure

This experiment had two conditions manipulated through the use of retirement during the learning phase: Random Interactive Retire (RIR) and Random Interactive No-Retire (RIN). The effect of these different modules on the performance of participants on the post-tests was measured in terms of accuracy and response time, measuring recall ability and fluency of the material. Both of the conditions consisted of a pre-test, a training phase, a post-test, and a delayed post-test identical in format to those in the previous experiments. During the training phase, participants completed a learning module dependent on their condition. Both conditions saw the countries presented in an active format without sequencing (for information on active format and sequencing, see Experiment 1). In the training phase RIR condition, items could be retired from the learning set based on the criteria described in Experiment 1. There was no retirement in the training phase of the RIN condition. In both conditions participants were trained until they reached certain learning criterion, the same as in the previous two experiments.

Results

Accuracy

Figure 1 presents the average accuracy of participants’ responses on the pre-test, post-test, and delayed post-test. From the pattern of results shown in Figure 1, both conditions appear to do equally well on the pretest, but on the post-test and delayed post-test participants in the Random Interactive Retire condition appear to do better than participants in the Random Interactive No-Retire condition.

The data were analyzed using a 2 (conditions: random order interactive retire, random interactive no-retire) x 3 (phases: pre-test, post-test, delayed post-test) repeated measures ANOVA test to compare the effect of retirement on the accuracy of responses. Mauchly’s test
indicated the assumption of sphericity had been violated $\chi^2(2) = 6.48, p = 0.04$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.86$). This revealed a significant main effect of phase ($F(1.72, 65.48) = 595.08, p < 0.001, \eta^2 = 0.94$), a main effect of condition ($F(1, 38) = 6.50, p > 0.02, \eta^2 = 0.15$), and a marginally significant interaction between phase and condition ($F(1.72, 65.48) = 3.22, p = .05, \eta^2 = 0.08$).

Participants improved significantly from pre-test ($M = 0.06, SD = 0.04$) to post-test ($M = 0.92, SD = 0.16$), paired-$t (39) = -34.08, p < 0.001, d = -7.37$. Participants improved significantly from pre-test to delayed post-test ($M = 0.78, SD = 0.15$), paired-$t (39) = -28.73, p < 0.001, d = -6.56$. There was a significant drop in accuracy from post-test to delayed post-test, paired-$t (39) = 4.62, p < 0.001, d = 0.90$. The accuracy of responses in the random interactive retire condition ($M = 0.56, SD = .08$) was significantly lower than in the random interactive no-retire condition ($M = 0.62, SD = 0.08$), ($F(1, 38) = 6.50, p > 0.02, \eta^2 = 0.15$). At the pre-test and post-test, the conditions’ accuracies were not significantly different ($p > 0.05$), but at the delayed post-test random interactive no-retire was more accurate ($M = 0.84, SD= 0.09$) than random interactive retire ($M = 0.71, SD = 0.17$), ($t(38) = -3.03, p = 0.004, d = 0.96$). This suggests that retirement might not be beneficial to accuracy of recall of factual information. To further examine these results, Experiment 4 was designed.

**Response Time**

Figure 2 presents the average response time of participants’ accurate responses on the post-test and delayed post-test. Because we were only interested in response time on accurate responses, we did not include pre-test response times because the accuracy at pre-test was too low across all conditions. From the pattern of results shown in Figure 2, participants in the
Random Interactive Retire condition appear to respond more slowly than participants in the Random Interactive No-Retire condition on the post- and delayed post-test.

The data were analyzed using a 2 (conditions: random order interactive retire, random order interactive no-retire) x 3 (phases: pre-test, post-test, delayed post-test) repeated measures ANOVA test to compare the effect of retirement on response times of correct responses. Mauchly’s test indicated the assumption of sphericity had been violated $\chi^2(2) = 24.94, p < 0.001$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.67$). The results show a significant main effect of condition ($F(1,37) = 6.56, p = 0.02, \eta^2 = 0.15$) but no main effect of phase ($F(1.33,49.33) = 1.74, p > 0.05, \eta^2 = 0.05$) or interaction between condition and phase ($F(1.33,49.33) = 0.89, p > 0.05, \eta^2 = 0.02$). The response time in the random interactive retire condition ($M = 4.09, SD = 1.01$) was significantly higher than in the random interactive no-retire condition ($M = 3.17, SD = 0.97$), ($F(1, 37) = 6.56, p = 0.02, \eta^2 = 0.15$). This appears to suggest that retirement results in slower response times on correct answers. To further examine these results, Experiment 4 was created.

**Experiment 4**

In the previous experiments, some conditions saw a significantly larger number of trials during the learning phase. Experiment 4 was done to determine whether this extra practice influenced the pattern of results observed in the Experiments 1-4.

**Method**

**Participants**

In the experiment, 40 undergraduate students (age range 18-34) were recruited from the University of California, Los Angeles. The initial 17 participants participated between 2007 and
2008, and an additional 3 participants were run in 2014. Participants received course credit as compensation.

**Apparatus and Stimuli**

The stimuli in this experiment were identical to that in Experiment 1.

**Design and Procedure**

This experiment had two conditions manipulated through the number of trials the participants saw during the learning phase: Random Interactive No-Retire (RIN) and Fixed Interactive No-Retire (FIN). The effect of these different modules on the performance of participants on the post tests was measured in terms of accuracy and response time, measuring recall ability and fluency of the material. Both of the conditions consisted of a pre-test, a training phase, a post-test, and a delayed post-test. The pre-, post-, and delayed post-tests were identical in format to those in the previous experiments. During the training phase, participants completed a learning module dependent on their condition. Both conditions saw the countries presented in an active format without retirement (for a more information about active format and retirement, see Experiment 1). In the RIN condition, participants were trained until they reach a certain learning criterion based on accuracy and response time. In the FIN condition, the phase always ended at 192 trials regardless of whether or not the participant had reached learning criteria.

**Results**

**Accuracy**

Figure 1 presents the average accuracy of participants’ responses on the pre-test, post-test, and delayed post-test. From the pattern of results shown in Figure 1, both conditions appear to do equally well on the pretest, but on the post-test and delayed post-test participants in the
Random Interactive No-Retire condition appear to do better than participants in the Fixed Interactive No-Retire condition.

We analyzed the data using a 2 (conditions: fixed interactive no-retire, random interactive no-retire) x 3 (phases: pre-test, post-test, delayed post-test) mixed analysis of variance (ANOVA) to confirm the effect of extra practice on the accuracy of responses. This revealed a significant main effect of condition ($F(1, 38) = 23.23, p < 0.001, \eta^2 = 0.38$), main effect of phase ($F(2, 76) = 343.39, p < 0.001, \eta^2 = 0.90$), and condition x phase interaction ($F(2, 76) = 13.20, p < 0.001, \eta^2 = 0.26$).

The accuracy of responses on the random interactive no-retire condition ($M = .62, SD = .07$) was significantly higher than on the fixed interactive no-retire condition ($M = .43, SD = .16$), $F(1, 38) = 23.23, p < 0.001, \eta^2 = 0.38$. Participants improved significantly from pre-test ($M = 0.06, SD = 0.05$) to post-test ($M = 0.82, SD = 0.26$), (paired-t (39) = -18.70, $p < 0.001$, $d = -4.06$), and from pre-test to delayed post-test ($M = 0.69, SD = 0.23$), paired- (39) = -17.48, $p < 0.001$, $d = -3.79$. There was a statistically significant drop in accuracy from post-test to delayed post-test, paired-t(39) = 4.39, $p < 0.001$, $d = 0.53$. Participants’ accuracy of responses on the post-test and delayed post-test differed significantly in the fixed interactive no-retire versus a random interactive no-retire condition. At pre-test there was no difference in accuracy of responses between conditions, $p > 0.05$. There was a significant difference in the post-test scores for participants who saw a large number of trials (i.e., in the random interactive no-retire condition, $M = 0.95, SD = 0.20$) and participants who saw a smaller number of trials (i.e., in the fixed interactive no-retire condition, $M = 0.70, SD = 0.25$), $t (38) = 3.63, p = 0.001$, $d = 1.10$, as well a significant difference in the delayed post-test scores for participants who saw a large number of trials ($M = 0.84, SD = 0.09$) and participants who saw a smaller number of trials ($M = 0.70, SD = 0.25$).
0.54, SD = 0.23), t (38) = 5.42, p < 0.001, d = 1.72, suggesting that the amount of extra practice on the random interactive no-retire improved the accuracy of responses on both post-test and delayed post-test as compared to those in the fixed order interactive no-retire condition who saw fewer trials.

**Response Time**

Figure 2 presents the average response time of participants’ accurate responses on the post-test and delayed post-test. From the pattern of results shown in Figure 2, both conditions appear to have similar response times at the delayed post-test, but on the post-test participants in the Random Interactive No-Retire condition appear to respond more slowly than participants in the Fixed Interactive No-Retire condition. Because we were only interested in response time on accurate responses, we did not include pre-test response times because the accuracy at pre-test was too low across all conditions.

We conducted a 2 (conditions: fixed interactive no-retire, random interactive no-retire) x 3 (phases: pre-test, post-test, delayed post-test) mixed analysis of variance (ANOVA) to examine the effect of extra practice on the response times of correct responses. Mauchly’s test of sphericity indicated the assumption of sphericity had been violated for the within-subject effect of phase on response times, $\chi^2(2) = 26.93, p < 0.001$, suggesting that the variances of the differences in response times between phases are not equal. To reduce the increase in Type I error rate in the following analyses with response times, we report statistical significance with corrected degrees of freedom using Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.64$). The results show no significant main effect of phase ($F(1.28,43.65) = 0.85, p > 0.05, \eta^2 = .02$), no significant main effect of condition ($F(1,34) = 0.54, p > 0.05, \eta^2 = 0.02$), and no significant interaction between condition and phase ($F(1.28,43.65) = 0.93, p > 0.05, \eta^2 = 0.03$). The amount
of extra practice on the random order interactive no-retire did not improve the fluency of recall on either post-test or delayed post-test as compared to those in the fixed order interactive no-retire condition who saw fewer trials.

**General Discussion and Conclusion**

In four experiments, we studied how interactivity, adaptive sequencing, retirement, and practice affect long-term retention of facts and the fluency and efficiency of fact recognition. The studies indicate no effect of sequencing, but do show an effect of interactivity. Interactivity seems to improve the accuracy of recall of factual information, but not the fluency of this recall. While initially it appears that retirement has an effect on long-term recall of factual information and efficiency such that randomly ordered stimuli are better learned than sequenced stimuli, this unexpected result can be explained in Experiment 4 by the uneven number of trials in the learning phases. Because those in the no-retirement condition saw nearly twice as many trials as the retirement condition (476 trials compared to 254 trials), the superior accuracy and response time can be explained by the participant having more practice with the items. In fact, when efficiency is examined (post-test accuracy gains divided by the number of learning trials invested) the non-retirement group is far less efficient than the retire group on both pre- to post-test efficiency and pre- to delayed post-test efficiency, as shown in Figure 3. When we compare efficiency across all conditions, we find that the passive/active condition is far more efficient than the others, suggesting that though pure interactive modules are superior to pure passive, a combined approach is the most effective. When learners are initially exposed to material before beginning to actively learn it themselves, they appear to learn the material more quickly. Overall the results support the active-learning hypothesis and constructivist models of learning that predict learning in an interactive setting increases learning through engaging the learners and
making them actively participate in the learning process (e.g. Moreno et al., 2001; Mayer, 1999; Coakley & Wolvin, 1997).

The results found in the experiments in this study do not support the findings of Mettler and Kellman (2014) or Mettler et al. (2011), however this could potentially be explained by the stringency of learning criteria used in the modules. Mettler set stricter learning criterion in his 2014 study (learner had to correctly identify a country five times out of six in under 3 seconds per item) than we did in the present four experiments (learner had to correctly identify a country three times on a row in under 10 seconds per item). The less stringent learning criterion in the four experiments here could have resulted in participants reaching learning criterion and ending the learning phase before truly mastering the material.

The effects of interactivity and adaptive sequencing on computer-based learning have important implications for the design of future programs. Currently, many e-learning systems are non-interactive and non-adaptive, but the outcome of this study and of those like it could provide compelling reasons for this to change. If teachers and program designers want to design effective and engaging learning programs for students, they must rely on the outcomes of such experiments to guide the design of their programs.
References


Figure 1. Average accuracy of country identification on pre-, post-, and delayed post-tests. The measure of accuracy is on a ratio scale from 0 to 1 where 0 means the participants identified none of the countries correctly and 1 means they correctly identified every country.

Figure 2. Average response time of correct country identification on pre-, post-, and delayed post-tests. Response time is measured in seconds.
**Figure 3.** Average pre-test to post-test and pre-test to delayed post-test change in efficiency. Efficiency is calculated by the gains in average accuracy on the pre-test to the post-test (or delayed post-test) each condition, divided by the average number of learning trials in each condition.

**Figure 4.** On the left is an example of the stimuli in the training phase of the interactive condition. On the right is an example of the stimuli in the training phase of the passive condition.
Neuropsychological Disorders and Memory

Mariam Hovhannisyan

Castel, Lee, Humphreys and Moore (2010) studied how children with and without Attention-deficit/hyperactivity disorder (ADHD) differ in their use of working memory. Children with ADHD possess several deficits, one of which is associated with the cognitive disability to use memory over time (O’Neill & Douglas, 1996). They hypothesized that children with ADHD would have a specific deficit in recalling and giving attention to the high-value items. Furthermore, Castel et al. presented children with words that were assigned a point-value of either high or low (between 1-12). There were eight lists shown to the participant and after each list they recalled as many items as they could. The selectivity index gave an accurate measure of sensitivity to the values of the word, such that it measured if a child successfully recalled a word based on the point value of the word and not just on how many words the child recalled. They found that children with the ADHD Combined type, which is characterized by both inattention-disorganization as well as hyperactivity impulsivity, were significantly less selective than the control group as well as the ADHD Inattentive type which is characterized by only inattention-disorganization. All of the children were able to recall more high value items than low value items, but children with ADHD Combined type were not able to efficiently maximize their score relative to the ADHD Inattentive type and the control group (Castel et al. 2010). This means that individuals with the ADHD Combined type were not able to increase their memory performance as shown by the selectivity index, thus leading to the conclusion that they have deficits in their working memory and are not able to efficiently recall high-value items.

A similar study by Castel, Balota, and McCabe (2010) looked at how the efficient use of
memory is influenced by aging and Alzheimer’s disease (AD). Just like the study above, a selectivity task was used in which items were worth point values. They hypothesized that the ability to selectively focus on information that differs in value is influenced by aging and Alzheimer’s disease. In short, they wanted to see if older adults give weight to important information (high-value items) that necessarily need to be remembered, compared to younger adults, and how that in the long run, can be used as a tool to detect early signs of AD. The selectivity task measured the efficiency of memory, not just the number of words recalled, and found that healthy older adults and young adults were able to significantly increase their selectivity index compared to adults with AD, who showed a lower SI. This means that young adults and healthy older adults improved their SI after each test, while the AD group did not. Although they recalled high-value items more than low-value items, relative to healthy older adults, adults with AD were unable to maximize their SI.

Both articles present very interesting topics related to memory and its efficiency in relation to selectivity of important information. Both articles were very similar in the sense that they conducted similar experiments, but in relation to different neuropsychological disorders. It was shown that most people overall recall more high-value items than low-value items, but people with neuropsychological disorders such as Alzheimer’s and ADHD do not maximize their score throughout the test. Although they did recall high-value items more than low-value items, their SI was significantly lower when compared to healthy older adults and younger adults. It would be interesting to see if children with ADHD and older adults with Alzheimer’s have these memory deficits in the same or around the same area in the brain. Hence, if it is around the same area in the brain, a relation may be formed between the two disorders and its implications on early diagnosis can be modified.
Comparing older adults with Alzheimer’s and older adults with ADHD may also give more of an insight on whether or not the disabilities and the location of those disabilities in the brain affect their memory in a similar way. Moreover, if it is similar, measures can be created to detect the early onset and diagnosis of either disorder. It would also be interesting to see how forgetting occurs in people with ADHD and AD, as well as in people with similar neuropsychological disorders. If value is put to certain items that need to be forgotten, do older adults, younger adults, and people with ADHD and AD perform similarly? This would be interesting to see because if people with neuropsychological disorders perform similarly on the value-directed remembering task, as well as a forgetting task, then this could possibly tell us if memory efficiency is related to long-term memory, such that the if the selectivity index increases, it will thus lead to remembering those items for a longer period of time.

The selectivity index, which measures the efficiency of memory, can also be used in relation to other topics. The implications of using this method of assigning value to items can be studied outside the range of neuropsychological disorders. For example, can selective attention to high-value words be used in a learning context? Can students allocate values to what’s more important in what they need to study? In learning, it would be interesting to see if people score better on tests depending on whether what they are studying is important information (high-value) or not. Furthermore, this could eventually help teachers in K-12 implement better curriculum for their students. Also, it is interesting to note, as a result of my observation, males and females tend to perform differently on the recall test. This could also be a factor in the studies mentioned above, such that the ability to recall high-values items can also depend on the gender of the person with the disorder.

During my time in the lab I’ve observed many interesting behaviors that people engaged
in during the experiment. For example, several of the participants would look away from the screen when a low-value word appeared. Perhaps people employed this behavior, because even if they didn’t want to encode the word, they would try to encode it anyways if they did look at the screen. I found this behavior quite interesting and believed that it might influence what items people encoded, such that if someone wanted to remember only high-value items, this may affect their later recall of a high-value item that came right after a low-value item. Additionally, perhaps trying to ignore the word but accidently looking at it may cause one to remember it better. I also observed, although I’m sure statistical analysis would be needed to confirm this, is that both males and females improved their score throughout the experiment, but males tended to score higher than females after each test. Also, males tended to recall more words and increase their score after being told what their score was. For example, if males had a low score in the first list many of them would increase their score on the second list. Hence, if I told them their score was low they tended to maximize their score much more rapidly than females. From my observations and looking at the raw data, it seemed that females increased their score at a lower rate than males.

My experience in the lab thus far has been very educational and interesting. I have learned a lot about memory and human behavior while working in the lab. For example, when running people for the experiment in which they had to recall words after every list, some participants would say the words out loud while encoding them and others would look away if the word had a low-value. The behaviors that I observed were fascinating to say the least; to see how different types of people learn the same thing. I also enjoyed the independency that was given in the lab; for me it promoted an active learning environment where I was able to increase my knowledge of memory as well as human behavior. Being able to work with other research
assistants as well as learning new things about memory and human behavior not only from the experiment, but also from my graduate advisor, was very valuable to me. I thought the lab meetings were also very informative and helpful. Seeing what graduate students were working on and how one topic such as memory can be studied alongside various other topics made me realize the value of research.

My only concern is that there isn’t a lot of involvement with the research assistants in the labs overall. Though I did get to meet other graduate students and speak with them, I believe a biweekly activity with the research assistants in all of the related labs would help promote discussion about the studies and topics within the various labs. It would also help the research assistants figure out what their specific interests are within the domain of psychology.

I hope to eventually propose my own study based on my work and observations in the lab. Human behavior, memory, decision making, judgment, thinking, and their relation to law has always been a special interest of mine. Moreover, the experience that I have gained while working in the lab cannot be replaced by classroom instruction. My experience has actually helped further my knowledge and expanded on the teachings in the classroom to real-world situations. The hands on experience I have gained better enables me to understand what I have learned in the classroom, as well as see the implications of research in a real-world context.
References


Major depression disorder (MDD) is largely a disease of attention. Individuals with MDD have a tendency towards depression, a compulsive inward turning of attention towards the symptoms of one’s own distress. Attention can be thought of in terms of two systems (neural activation networks), the default mode network (DMN) and the task positive system (TPS/TPN). The DMN can be thought of as the network activated when the mind is at rest and withdrawn from cognitively demanding tasks or social interaction. The DMN is where our mind goes when it is not dealing with external environmental demands and is allowed to reflect upon itself. The TPN can generally be thought of as the network which serves an opposite function than the DMN, it is the network activated when attention is shifted from the self towards the external world or the solving of complex problems.

One popular theory of MDD suggests that those who suffer from MDD have a “sticky” DMN. Those with a sticky DMN have difficulty shifting suppressing the activity of the DMN during cognitively demanding tasks and have difficulty recruiting the FPTCN, a part of the TPN. This difficulty suppressing the DMN causes those with MDD to have difficulty shifting attention from internally generated states to externally oriented tasks. One manifestation of this is rumination by those with MDD while trying to solve complex problems. This rumination tends to demand cognitive resources that would be better directed towards task completion and can result in dramatically impaired performance on tasks.

Bergman
Bergman (2014) noted that connectivity patterns that related to subject mood were strikingly difficulty for those with MDD and those without mental health issues. This lead the researchers to believe that those with MDD and those without had different mood regulating mechanism while experiencing unconstrained thinking and induced rumination.

A widespread belief is that those with MDD are consistently ruminating, even when their mind is at rest, however the findings of Bergman clash with this popular notion. Not only were measures of mood significantly different when subjects were allowed to let their minds wander but neural activation patterns also shifted significantly. Like the subjects with MDD, healthy participants had a lower mood when in a state of induced rumination; however, their neural activation pattern differed. Healthy participants had consistently widespread neural activation during both open contemplation and induced rumination while participants with MDD had exacerbated activation in the DMN while in a state of induced connectivity. It appears that participants with MDD are able to enter a state of rumination with less effort, as if it is a state in which they more naturally effort, as if out of habit. Despite the ease with which those with MDD enter a ruminative state, the contention that they are by default in a ruminative state seems unlikely.

Based upon these findings, Bergman suggests that to treat those with MDD increasing overall connectivity while reducing the role of the DMN may be effective in reducing the distressing rumination that is characteristic of depression. Overall, Bergman’s findings seem to back up what has been hypothesized for some time. Those with depression when compare to healthy participants more easily slide into ruminative states and have fundamentally different mood regulating mechanisms. Counter to popular belief, rumination is not the default state of free thought for those with depression who actually show very similar normal activation patterns
but it is simply requires less effort for those with depression to enter and sustain a ruminative state.

Hamilton

Hamilton (2012) provides a review of the approaches to understanding MDD from a neural systems approach and more specifically comments on the role of the default-mode, executive, and salience in the disorder. To Hamilton, the brain can be parsed into the DMN, active when the mind is at rest, and the TPN, the network of structures that are active during “performance of attention-demanding tasks”. These two systems of mind are in competition with one another for attentional resources and have inversely correlated activity levels. The TPN can be further divided into the executive network (EN), positively related with executive task performance, and the salience network (SN), related with ratings of state anxiety.

These three major networks each contribute uniquely to the symptoms of MDD, as such their independent functioning and interrelatedness are important to understanding the basis of MDD. Rumination is closely related to the activity of the DMN, emotional disinhibition to the EN, and emotional over-reactivity to the SN. Rumination patterns in the DMN are thought to be vital to maintenance of MDD. Those with MDD have trouble deactivating the DMN when exposed to self-relevant negative stimuli and it appears that when possible, they reflect upon stimuli in terms of how it relates to themselves, rather than remaining passively observant.

One structure that has been considered particularly important in switching between the DMN and the TPN is the right anterior insula, a component of both networks. It
appears that the right insula engages the TPN when the DMN is performing at heightened levels and seems to play a role in switching between dominant activities in the two networks. Major depression is more than a simple regression into the ruminative thought patterns in the DMN, rather it is a more sophisticated interplay between the DMN, EN, and SN. Each network plays its own role in the maintenance and expression of depression symptoms.

Mittner:

Mittner (2014) suggest that mind-wandering is a ubiquitous part of everyday life that is characterized by “inefficiencies in executive control processes”. Mind-wandering is an activity reflected in the functioning of the DMN, particularly the PCC and mPFC. Activation levels of the DMN and ACN can reliably predict mind wandering. Mind wandering has a pronounced effect on executive processes controlling goal monitoring. SART has proven to be an effective tool for inducing mind wandering. Due to its boring and repetitive nature, participants are likely to lose concentration over time and show the effects of mind wandering: increased error rates, anticipatory responses, and response omissions. Although the SART is an effective tool for showing the effects of mind wandering, its highly simplistic nature makes it difficult to discern how other higher thinking processes may be affected by mind wandering.

Conclusion:
The role of several neural networks is being better understood through advances in neural modeling and imaging techniques. One of the most well documented networks in this realm is the DMN, which has been implicated for its role in rumination and mind wandering. Despite the common conception that patients with MDD fall into ruminative thought patterns frequently when they are given the chance to think freely, it appears that the case is not so simple. Although the mind may wander and the DMN is implicated in this process, unusually heightened activity that differs from the activity of healthy individuals is shown when induced rumination occurs, not free thinking. It also appears that depressed patients can more easily slip into states of heightened DMN activity and rumination than healthy patients, which supports the notion that rumination is a key factor in the development and maintenance of depression.

The symptoms of depression are not isolated to the DMN alone, rather depression appears to be a consequence of a dysfunction relationship between the TPN and DMN. The TPN shows involvement in both emotional disinhibition and overreactivity characteristic of depression. The right anterior insula, an area involved in both the TPN and DMN, appears to play a role in shifting between the two activation patterns. Further research into the role of this region and how it functions when participants are completing a higher level cognitive task (like the N-back) could help shed some light on the complex interplay between the networks and how the dysfunction between them leads to difficulty for those afflicted by MDD on demanding cognitive tasks.
Sources


Social Working Memory
Natalie Saragosa-Harris

Introduction

Previous research has focused on nonsocial cognition, yielding many reports of the brain regions involved in working memory and the effects of cognitive load. While processing nonsocial environmental cues such as shapes or spatial relations is important to everyday life, it is hypothesized that what sets humans apart from other species is the ability to think and interact socially. According to the social brain hypothesis, primates evolved large brains in order to regulate social information and maintain complex social systems (Dunbar, 1998). This idea has inspired researchers to study the brain mechanisms that allow people to think socially and to manipulate multiple pieces of social information. Current research focuses on social cognition and provides a foundation for future research in social working memory training and possible clinical applications (Meyer, Spunt, Berkman, Taylor, & Lieberman, 2012).

Social Load and fMRI

It is still unclear exactly which mechanisms are associated with maintaining numerous pieces of social information. The ability to manipulate multiple pieces of social information, or “social load”, is integral to social working memory. A recent study conducted by Meyer et al. (2012) examined networks in the brain that are engaged when processing social information, specifically regions that are sensitive to social load. Researchers developed a task in which participants were presented with the names of their friends and told to mentally rank the friends in terms of a given trait, such as “funny”. Subjects would then answer a true or false question
about their ranking. The paradigm varied the social load, or the amount of social information to maintain, on each trial, as subjects were either presented with two, three, or four names. Subjects completed these tasks during an fMRI scanning and researchers examined parametric increases in brain activity as a result of increasing levels of social load (Meyer et al., 2012). The study also addressed whether mechanisms involved in social working memory and social cognitive ability could be related. Given that lateral frontoparietal activity in nonsocial tasks is connected to general fluid intelligence (Conway, Kane, & Engle, 2003), researchers questioned whether this could be true for social tasks and respective brain regions as well.

In the results, as social load increased, researchers found increased activity in the dorsolateral prefrontal cortex, superior parietal lobule, and SMA, all of which are regions involved in working memory. This is consistent with previous findings about working memory and nonsocial tasks. As load increases, activity in these regions does as well. In addition, activity increased in the dorsomedial prefrontal cortex, anterior paracingulate cortex, TPJ, and precuneus/posterior cingulate cortex. These areas are associated with “mentalizing”, the process of considering the mental states or characteristics of other people (Meyer et al., 2012).

The finding that these mentalizing regions increase parametrically with increased load differs from previous research on working memory. The mentalizing network is almost identical to the default-mode network, the region that is active when humans are at rest. That is, when not engaged in a effortful cognitive task, this region is found to be active. In the past, researchers found that when one system has increased activation, the other has decreased activation. During an effortful task, for example, researchers found that standard working memory system has increased activation during a task, while the default or mentalizing system typically has decreased activity (Raichle & Snyder, 2007).
This inverse relationship, however, was previously studied as a result of nonsocial cognitive demands. The effect of social load, in contrast, revealed linear increases in both the medial frontoparietal regions and the lateral frontoparietal system. Thus, both the network involved in social cognition and the network involved in standard working memory was engaged in these social tasks. Effortful processing, therefore, does not always guarantee a decrease in these default-mode regions; rather, activation of the network is domain-specific and is increased during effortful tasks that involve social load (Meyer et al., 2012).

Researchers also analyzed whether there exists a relationship between activation of certain networks and social cognitive ability. They measured social cognitive ability in terms of trait perspective taking. In doing so, they questioned whether the areas involved in social competence and reasoning were distinct from those involved in general intelligence. The data suggest that only regions associated with social cognition, specifically MPFC, are correlated with perspective-taking ability. Activation of traditional working memory regions was not correlated with trait differences in perspective-taking ability (Meyer et al., 2012). Individuals were more likely to show increases in this region as a result of load if they had greater trait perspective-taking. Furthermore, the findings directly correspond to the social brain hypothesis, as MPFC is the only part of the frontal cortex that is significantly larger in humans than in other primates (Semendeferi, Armstrong, Schleicher, Zilles, & Van Hoesen, 2001). Further research suggests that differences in MPFC size are correlated with social cognitive competence (Powell, Lewis, Dunbar, Garcia-Finana, & Roberts, 2010) and the size of one’s social network (Lewis, Rezaie, Brown, Roberts, & Dunbar, 2011).
Working Memory Training

The ability to reason and address new problems without relying on prior knowledge is essential to managing demanding environments and is associated with professional success. This capability, considered fluid intelligence, is related to working memory. Individual differences in working memory capacity and the ability to derive relations, for example, have been found to account for differences in various fluid intelligence tasks (Jaeggi, Buschkeuhl, Jonides, & Perrig, 2008). Jaeggi et al. (2008) conducted a study to question whether one could improve fluid intelligence by training working memory.

The study focused on training with a working memory task and transferring this to completely separate measures of fluid intelligence. Based on the idea that working memory and fluid intelligence both depend on processes in the lateral prefrontal and parietal cortices, researchers set out to transfer the effects of training memory to fluid intelligence tasks. In the task, participants were presented with two series of stimuli (either a series of letters or a series of spatial locations on the screen). They were asked to determine whether the stimulus presented was the same as the stimulus from a certain number of items previously. Subjects underwent between 8 to 19 training sessions in which they completed this working memory task. In addition, subjects were tested on a measure of fluid intelligence before and after training and their results were compared to the control group, which underwent no training (Jaeggi et al., 2008).

Researchers found a significant improvement in the fluid intelligence measures in the group that received training. Furthermore, they found that the increase in fluid intelligence was not due to preexisting, individual differences. That is, neither initial differences in fluid
intelligence nor working memory accounted for the change in fluid intelligence. Furthermore, it appears that the change was dosage-dependent; more working memory training resulted in greater fluid intelligence improvement (Jaeggi et al., 2008).

Clinical Applications and Current Research

Understanding the processes involved in social cognition and the relationship between different mechanisms provides essential information for assessing unique human capabilities. However, this research does not only provide insights from an evolutionary perspective. These findings could potentially be used in assessing conditions related to deficits in social cognition and working memory such as schizophrenia, autism, and social anxiety (Meyer et al., 2012). From diagnosis to treatment, understanding the complexity of social interactions and the effects of social load provides a foundation for future intervention. Drawing on Jaeggi’s study with working memory training, it is possible that similar techniques could be used in a social context.

This quarter, I worked on a project that addressed this idea. The project questioned whether social working memory training could improve social cognitive reasoning and ability in the same way working memory training improved fluid intelligence. This ongoing project, therefore, could potentially not only emphasize clinical applications but also be pertinent to everyday life and social ability.


Perceptual learning is improving one’s skills at discriminating between a set of stimuli according to some (often implicit) criteria. The abilities gained through perceptual learning can be as simple distinguishing two slightly different shades of the same color, or as difficult as arranging a group of paintings by artist or learning to read—that is, matching written words to spoken ones. In the past, techniques from perceptual learning have been applied to the acquisition of other types of knowledge, including category learning (Mettler & Kellman, 2014) and even mathematics (Kellman, Massey, & Son, 2010).

One facet of perceptual learning that has been identified is the distinction between discovery effects and fluency effects (Kellman, 2002). Discovery refers to finding features pertinent to a given classification. For example, a learner might discover that the shape of a letter is important in terms of matching it to a sound, but that its size is irrelevant in this regard. Fluency, on the other hand, is the automaticity of perceptual categorization. To return to our reading example, as a learner’s fluency with the written word increases, he or she not only becomes faster, but also devotes fewer attentional resources to the mapping between spoken and written word, freeing them up to focus on the higher-order concepts expressed in the text itself.

Although they have been identified theoretically, there is still little known about exactly how and when each component affects learning—or, indeed, if one of these is more crucial for some part of learning than the other. The difficulty arises when we attempt to distinguish the two, since both contribute to learning, and there are rich interactions between them as well. In
the “chunking” (unitization) phenomenon, for instance, the process of discovering how to split up higher information into chunks that can be manipulated in working memory can lead to more efficient, and therefore more fluent processing of a stimulus. The reverse may also occur, in which fluency in the recall of pre-existing knowledge can enable learners to use that knowledge during new learning, thus scaffolding their way towards discovery of other relevant perceptual factors. However, the general assumption in perceptual learning is that discovery is still the prerequisite of fluency, at least initially.

In the present experiment, we were particularly interested in adaptive learning. In its most basic form, “adaptive learning” refers to a study technique wherein items are chosen based on the learner’s previous performance (Mettler, Massey, & Kellman, 2011). Generally, the purpose of this technique is to focus attention on items that have a lower “learning strength”—i.e., items that are more likely to be forgotten later. Previous studies on the spacing effect have suggested that items should be presented again just before they are about to be forgotten. The algorithms used to select the next item in our implementation of adaptive learning are designed to estimate this optimal point for each individual learner. This model takes into account the learner’s accuracy during the training phase as well as his or her response time, which are intended to serve as rough measures of discovery and fluency, respectively.

For this particular experiment, we were also interested in studying adaptive learning in conjunction with variation in the amount of information presented as the “cue” for each item. More precisely, we decreased the number of representations used for each item during training. The goal of this manipulation was to encourage abstraction, the rationale being that the integration of multiple representations in the stimulus would create a more holistic, richer mental model of each item, rather than a mental image in which each item is simply paired with a single
arbitrary symbol. Previous research by Schnottz and Bannert (2003) has indicated this to be beneficial for learning.

Method

Participants

Thirty-one college-age participants took part in the study, 19 of whom were concurrently enrolled in a chemistry class at Collin College in Texas, and 12 of whom were recruited from the psychology study pool at the University of California, Los Angeles. The UCLA students received extra credit points in their psychology courses for their participation.

Materials

The perceptual learning module was intended to teach students about the bond angles and hybridization of 12 different chemistry molecules. Each question in the experiment was presented in multiple-choice format, and all questions were displayed and answered using a computer. Participants saw a cue consisting of one, two, or three different visual representations of each molecule, and were instructed to match the cue to what they thought was the correct answer. Participants used a mouse to indicate their choice.

The three types of visual representations used in the experiment were a three-dimensional computer illustration, a two-dimensional Lewis structure, and an empirical chemical formula. In the three-dimensional illustration (top left in Figure 1), each molecule was displayed as a computer graphic of a ball-and-stick model revolving around a central axis. The two-dimensional Lewis structure (bottom left in Figure 1) contained the chemical symbol of each atom in the molecule, with lines indicating bonds between the atoms and dots marking lone pairs of
electrons. Finally, the empirical formula (bottom right in Figure 1) contains only the chemical symbols of each atom, the relative numbers of each type of atom in the molecule, and the ionic charge of the molecule. These three representations were ranked according to perceptual salience, the three-dimensional ball-and-stick model being the most salient, and the empirical formula least salient. At the start of the training phase, all three representations were present in each cue. After each learner reached a certain criterion on a given item (4 presentations correct in a row, with each response under 20 seconds), one of the representations was removed from the cue for that item; first, the ball-and-stick model was removed, then the Lewis diagram—leaving only the empirical formula in the last part of the training phase. When only the empirical formula was left in the set, items dropped out of the pool of available questions entirely when the aforementioned criterion was met.
Figure 1. Question from the beginning of the training phase, containing all three representations. As the learner progressed through the training phase, the cue for each question contained fewer and fewer representations.

**Conditions**

This study featured a between-participants design, in which each participant was assigned randomly to either one of three scheduling conditions. In the first condition, the order of questions was chosen adaptively, based on each students’ performance and response time on each item throughout the training phase. In the other two conditions, instead of matching the order of study items to fit each participant’s individual learning strength, the question selection algorithm in these conditions only took into account the number of items presented since the previous appearance of a given item, and chose the next item such that the intervals between each presentation of an item would continuously increase at a roughly constant rate.

The two expanding conditions differed in the scheduling of items after a representation was removed from the cue. In one expanding condition (which we will refer to as “expanding with reset”) the interval between subsequent presentations of each item contracted to its initial value (one) when a representation was removed from the cue. Conversely, in the other expanding condition, the spacing between presentations expanded as before, irrespective of the change in the number of representations.

**Procedure**

The experiment took place in two sessions spaced a week apart. In the first session, all participants took a 20-item multiple choice pretest to assess their existing chemistry knowledge. A perceptual learning module was then administered, in which the method by which items were
scheduled varied according to the participant’s assigned condition. Following this training phase, participants performed a 20-item multiple choice immediate posttest. The delayed posttest one week later was also multiple choice, also 20 questions, and also picked from the same pool as the questions for the pretest and immediate posttest.

Results and Discussion

Before the results are discussed, we must mention a few limitations of the present study. The sample size was rather small, especially taking into account the rate of attrition. Only 20 of the participants completed the second session of the experiment; thus, only 20 data points were used to compare across the two sessions. Moreover, one possible confound was that some of the participants were simultaneously enrolled in a chemistry course. This means that these participants could have been exposed to the same information sometime between the first and second sessions of the experiment. In other words, these results should be taken with a grain of proverbial salt.

That being said, preliminary analyses of the data indicate a difference in the average retention from immediate to delayed posttest between the adaptive and combined expanding conditions. Specifically, in the adaptive condition, a paired-samples t-test revealed no significant difference ($t(4) = 1.02, p = .36$) between immediate posttest accuracy ($M = 0.52, SD = .24$) and delayed posttest accuracy ($M = 0.60, SD = 0.23$) scores. In contrast, a paired-samples t-test on the expanding conditions demonstrated that the delayed posttest accuracy score ($M = 0.51, SD = 0.13$), for this condition was significantly lower ($t(14) = 2.29, p = .038$) than the immediate posttest accuracy score ($M = 0.59, SD = 0.16$). An unpaired t-test showed no significant difference ($t(18) = 0.77, p = .45$) between immediate posttest accuracy scores in the adaptive
condition \((M = 0.52, SD = .24)\) and the expanding condition \((M = 0.59, SD = 0.16)\). These trends suggest that the adaptive condition was better at promoting long-term retention, even though there was no benefit over the expanding condition in the short term.

Seen through the lens of discovery and fluency effects, we must ask whether the discrepancy in retention between scheduling conditions is due to variation in discovery, fluency, or both. From looking at the average response times (in ms) from the training trials, it seems as though there is no significant difference \((t(29) = 0.04, p = .97)\) between the adaptive condition \((M = 5256, SD = 2570)\) and the expanding conditions \((M = 5222, SD = 1640)\). This suggests that any advantage of one sequencing condition over another lies not in improved fluency, but instead in the realm of discovery. However, in order to ascertain whether this is the case, we must develop a measure of discovery that is more fine-grained than simply looking at whether a given question was answered correctly or incorrectly.

In this study, the distractors for the multiple choice questions were chosen arbitrarily, and not varied across the training trials. Future studies could use tailored distractors in order to ascertain which features participants are judging as important for classification, and thus monitoring how the relevant features change over time. This improved measure of discovery could help us determine the points at which discovery is taking place. The hope is that this will enable us to see exactly how scheduling interacts with discovery in order to create a better model of memory, as well better learning systems that can encourage discovery precisely when it is most crucial for learning—whether through adaptive scheduling or some other manipulation.
Bibliography


Mean Diffusivity and the Microstructural Properties Underlying the Hippocampus

Omri Raccah

This paper will be focused on investigating the benefits of using mean diffusivity (MD) to understand the physical and architectural attributes of the hippocampus. In doing so, it is possible to make inferences about a wide-range of clinical disorders with respect to detailed anatomical differences in hippocampal structure. This provides the potential to predict and assess Alzheimer’s disease, epilepsy, and schizophrenia, among other disorders. In addition, examining connectivity through lends of mean diffusivity allows for observing hippocampus networks with relation to other cortical areas. This elucidates brain regions that are associated with memory consolidation and other hippocampal functions. Furthermore, MD can serve as a powerful tool for investigating individual performance differences on hippocampus-related tasks.

Diffusion tensor imaging (DTI) has had extraordinary success in understanding the microstructural correlates of neurological disorders. DTI is a magnetic resonance imaging (MRI) method that measures diffusion of water molecules in biological brain tissue. In this way, DTI is able to provide information about the structure and geometric orientation of neural matter on a voxel-wise basis. Aside from clinical applications, DTI is utilized to visualize neurological connections between cortical areas. This is useful for interpreting activation across regions and finding interconnected neural networks. Additional research is focused on observing enhanced cognitive abilities as a function of particular structural properties.

An article by Bihad, Le D. et al. (2001) explains two measures obtained from DTI data to provide information regarding the microscopic structure of neural tissue. Mentioned previously,
MD represents the average-squared diffusion and obstructions to diffusion. The authors note that this interpretation is similar in white and grey matter respectively. In addition to MD, fractional anisotropy (FA) is a measure of the degree of anisotropy, or level of directional dependence. This relates to the directionality of diffusion and provides a method for understanding the structural architecture of myelinated fibers. These measures are often used together in order to provide a detailed interpretation of a given phenomena (Alexander et al, 2007).

The most common application of mean diffusivity in the hippocampus is identifying structural correlates of clinical disorders. This area contributes to predicting the onset of dementia through examining hippocampal integrity. By doing so, this measure has the potential to provide earlier diagnoses of various neurological disorders. In addition, assessing the disease state of a patient is widely studies with relation to hippocampal structure and volume.

As mentioned, research in this area is often focused on the predictive attributes of hippocampal microstructure. These studies use mean diffusivity in order to identify these subtle physiological changes. A paper by Fellgiebel, A. et al. (2006) explores this relationship in patients diagnosed with amnestic mild cognitive impairment (MCI). MCI is a disorder impacts brain function and is often present prior to the onset of dementia. The study consisted of 13 MCI patients assessed over a 19-month period. MD and FA values were collected for the hippocampus of each patient. During this time interval, 6 of 13 patients had converted from MCI to dementia. Higher rates of MD were observed in the left hippocampus of the patients diagnosed with dementia. The authors attributed this effect to a loss of neural tissue allowing for an elevated degree of diffusion. Furthermore, it was noted that FA and volume measures were not sufficient in detecting these changes. Other studies found similar effects in MCI patients with
respect to MD in the left hippocampus (Muller et al, 2007). This area of research elucidates potential applications of hippocampal MD of MCI patients in identifying early signs of dementia.

Using MRI, hippocampal volume has been shown to decrease in patients’ diagnosed with schizophrenia (SZ). MD provides a method for investigating these structural effects on a microscopic scale. Chiappolini, C. et al. (2014) focused on explaining age-related effects on microstructure of the hippocampus in SZ patients. In the study, MD was obtained for 85 SZ patients and 85 healthy control subjects. Age was shown to correlate significantly with higher MD in the right and left hippocampus. This effect was present respectively for both healthy controls and SZ patients. The authors concluded that the reduction in hippocampal integrity could not be accounted for during a specific time frame. Rather that structural decay occurs gradually and is related to aging. Furthermore, SZ subjects showed earlier organizational abnormalities and a more rapid decline. In the study, MD served to identify structural abnormalities in order to consider deterioration of hippocampal tissue. The fornix, a central output region connecting the hippocampus to other cortical areas, has also been studied using DTI in SZ patients. Significant increases in fornix MD has been shown in SZ and suggested to relate to hippocampal properties (Kuroki et al, 2006).

Hippocampal MD has been examined with respect to other neurological diseases. For instance, a study by Weishmann, C. et al. (1999) focused on investigating hippocampal MD in patients diagnosed with refractory epilepsy. Epilepsy is often accompanied with hippocampus sclerosis (HS). This condition is categorized by structural abnormalities in the hippocampus. Thus, the authors were focused on assessing diffusion in this area. In the study, 14 patients and 6 controls were scanned and measures of FA/MD measures were generated in the hippocampus. The authors found a negative correlation between hippocampal volume and MD across
participants. However, the others did not find significant differences in FA measures between patients and healthy controls.

Patients with Parkinson’s disease (PD) have also been studies with regard to higher measures of MD. A paper by Kim, J. et al. (2013) examined MD in healthy controls and PD patients across white matter and deep grey matter areas in order to identify subtle structural differences. The authors found higher levels of MD among PD patients in a wide-range of motor and non-motor regions. Higher levels of MD were also noted to be present in the hippocampus and fornix. Additionally, The authors noted that differences FA between PD patients and healthy controls were not present. This study demonstrates that MD can serve as a robust measure for microstructural disparities across regions.

Studies are often concerned with identifying individual differences in memory performance with respect to diffusion in the hippocampus. Along these lines, DTI is used to explain the effect of microstructural integrity on various memory-related tasks. For instance, a paper by Heijer, T., et al (2012) investigates the relation of verbal memory performance and hippocampal MD in healthy participants. The purpose of this study is directed at examining whether DTI can be used to predict the onset of Alzheimer’s disease, in contrast to using measures of hippocampal volume. The participants chosen for the study (n=892) were healthy individuals older than 54 years of age. Measures of MD, FA, and hippocampal volume were obtained for each subject respectively. Subjects were given a verbal memory test in order to evaluate memory performance. The test was administered in an MRI scanner and involved memorizing a list of 15 words and subsequently recalling as many as possible. Measures of MD in the right and left hippocampus showed a significant correlation with memory performance. That is, higher MD was found to be associated with fewer words being recalled across
participants. Additionally, the authors found no association between hippocampal volume and performance on the memory task.

MD in the hippocampus has showed to be advantageous in understanding the physical properties underlying various diseases and behavioral measures. Hippocampal MD has the potential to serve as a powerful tool for predicting signs of dementia. Additionally, this method can be used to assess disease states of patients diagnosed with SZ and Alzheimer’s. Further applications in epilepsy and Parkinson’s disease may enable researchers and clinical doctors to discern the microstructural correlates of such disorders. Studying MD in hippocampal related regions such as the fornix provide intuition into hippocampus associative areas. In this way, cortical areas related to hippocampal memory functions can be taken into account in fMRI experiments. Recently, studies have focused on correlating behavioral performance with regional MD measures across healthy participants. In the hippocampus, this area of research lends itself to understanding the relationship between subtle microstructural differences and performance on memory abilities.
References


Finding out I was going to be a part of the Learning and Forgetting Lab, I was fascinated with the research I was going to be involved with. The goal of the lab is to “promote learning and memory performance within educational contexts through the investigation of principles in cognitive psychology”. As a student I felt like I would be able to contribute my thoughts on effective studying strategies while learning strategies to implement myself for my personal education. I clearly remember meeting with my graduate mentor Erin Sparck and being intrigued by her focus on effective study strategies for students, especially the “Confidence weighted multiple choice test format” technique. Erin’s research focus in this lab is to improve student’s metacognitive awareness and learning to increase productive retrieval for tests. Her previous research includes figuring out whether confidence weighted multiple choice tests as a studying tool increases overall memory performance. Her previous results show that confidence weighted multiple choice tests can improve final test performance on non-tested but related information under some circumstances.

The testing effect is the observed phenomena in which the act of taking a test can influence and improve the retention of that knowledge and directly promotes learning(CogFog powerpoint). The testing effect has been observed for different formats including multiple choice and can increase memory performance scores up to 25%. The key to the improvement in the scores is the retrieval of information. The more you retrieve information, the stronger it becomes in memory thereby making it more easily retrievable the next time the information is needed. However, there is also a negative effect associated with repeated retrieval. The more repeatedly something is retrieved, the less other items of information, with a related base, are able to be
retrieved. Applying to pre-tests and actual tests students take, the goal was to see what happens to information that was to initially tested on (that did not appear in the pretest but does on the actual test). Teachers supply and make students take practice tests, just a sample of what is likely to be on the test. It is not always the chance that the exact questions on the practice test show up on the actual test. How does this affect performance scores? Typically, the activation of one semantic node will spread to other semantic nodes through its associative network (Cogfog powerpoint). However, this diminishes with repeated retrieval.

Experiment 1 calls for participants to read an article about a Toucan bird. Then half the participants were asked to answer questions based off cued recall (essentially pretested)(ex: A) Where do toucans sleep at night). While the rest read statements (“cheating condition”) (ex: Toucans sleep in tree holes at night.) Then they all participated in a cued- recall test after a 24 hour delay period. Information that was pretested had the highest probability of accuracy recall, non-tested but related information was recalled less but still more than the control condition who was not previously tested. These findings support the idea of the testing effect and benefits of pretests. The experiments were then applied to another testing format, multiple choice. In an experiment by Little, Bjork, Bjork, and Angello (2012), participants were asked to read one passage and either answer cued recall or multiple choice pre-test. Then the participants read another passage after which they were not tested. The final test, cued-recall, consisted of previously tested information, related information, and information that was not previously tested. The data reveals that previously tested questions elicited more accurate responses for both multiple choice and cued recall(Bjork et al 2012). Control questions resulted with about the same amount of correct responses for both multiple choice and cued recall questions. The results also showed that initial multiple choice test prompted enhanced recall of related information. If the
multiple choice options are equally plausible and competitive, multiple choice can trigger productive retrieval processes (Bjork et al. 2012). In a replicated experiment, half of the participants received feedback on the test to discover whether correcting erroneous beliefs are beneficial for final testing. Feedback had a positive relation with number of correct responses for previously tested information for both cued recall and multiple choice but not for related information. In general, multiple choice testing resulted with better final test performance for related information whether or not feedback was given. Cued-recall initial test showed no benefits of this and actually showed worse performance score. Considering that related information becomes easier to recall, there is a benefit for initial non test information under the multiple choice format specifically, not with cued-recall.

Erin’s research question was, “How can we get learners engaged in a broad strategy type of thinking without direct instruction. Is it possible to make a more effective multiple choice test that improves accuracy and recall?” Her research was based off a confidence weighted triangle multiple choice format known as Bruno’s IRT, as shown in figure 1. IRT may be effective as pretests for students as teachers should be able to see if there is a common pattern of misunderstood information among students and why. Students are able to express partial knowledge recall of information previously learned, which can guide teachers to use different, more effective teaching strategies. Being able to express this partial knowledge shows their representation of the information but does not punish the students merely for not choosing the one accurate multiple choice option. The modification of the IRT triangle prompts learners to make metacognitive judgements, shown in figure 2. The purpose of experiment 1 was to see if there was a benefit of the confidence weighted multiple choice for related information. The experiment consisted of 3 conditions: standard multiple choice, confidence weighted multiple
choice, and study only. The participants would read passages, take a pretest or play tetris, read another passage, have a tetris distractor task, and lastly take the final cued-recall test. The participants were also asked to answer a survey regarding the confidence weighted-triangle format for multiple choice to understand their opinions. Majority of the students liked the idea of partial credit, thought it assessed their knowledge in a better manner and made them recall and assess information about the answers more carefully. However, they also stated that they thought it did not help them on their final test. The results expressed that the number of highly confident errors did have a relation to the final test score. The final performance scores did not turn out being as high as hoped. It was believed that the relational thinking of competitive answer alternatives along with the confidence judgements that provided a benefit for recall.

Confidence can be highly supportive of one’s answer, but it can also be misleading and cause errors. Another research question regards whether feedback can influence recall for the final test. If people were given feedback after making highly confident errors, are they more likely to make those same errors or more likely to remember the corrective feedback. Feedback in general can either support one’s confidence or correct one’s confidence in regards to future recall. In Experiment 2, feedback was given to participants during the initial test (Ex: The correct answer is ____). The results show that the confidence weighted benefit actually goes away. The overall testing performance decreased with feedback. It is speculated that maybe participants stopped actively participating, expecting the correct answer to always show. Another experiment was conducted in which some participants were allowed to have a restudy period. Unfortunately, the results still show no confidence weighted benefit. The participants who were allowed to restudy did not score much higher than those with one study period.
The experiments involving the confidence weighted multiple choice reveal that only under some circumstances will this test format improve final test performance. It was also concluded that subjects were unaware that pre-testing is helpful for being later tested on related information (Cogfog powerpoint). This research about educational, studying strategies is essential to effective teaching, learning and studying. Other and future research should further study the effect of multiple choice, confidence, and feedback as well as different types of informational material and test presentations. The confidence weighted multiple choice format so far has not proven to be the most effective strategy for learning. However, the Bjork lab has identified several other effective learning strategies such as spaced intervals and interleaving. Continuing my involvement in the Learning & Forgetting lab will further allow me to analyze current research and possibly pursue my own research ideas about metacognition and studying.

References

Figure 1. Bruno’s IRT.

Figure 2. Modified Confidence Weighted Multiple Choice Triangle.
Using Photographic Life Logging to Explore Autobiographic Memory

Tim Yu

Introduction

Our ability to remember past events from our own lives is made possible by our autobiographical memory (AM), a recollection of personal semantic information (facts about the self, such as knowing what year we were born in) and personal episodic information (unique events and personal experiences, such as remembering the moment we got accepted to our dream college). Compared to the recalling of personal semantic information, the recalling of personal episodic information is a much more complex and constructive process, as it requires the search and integration of relevant information from a number of different subsystems (i.e. sensory information such as visuals, emotions, etc.), and the re-experiencing and recollecting of specific past events (Wheeler, Stuss & Tulving, 1997).

In the past, many studies have been done on memory with laboratory-based stimuli, otherwise known as laboratory memory (LM)—operating under the assumption that the use of micro-events could capture the core process of memory, and also sacrificing ecological validity (Cabeza et al., 2004; McDermott et al., 2009). On the other hand, studies that focus on autobiographical memory sacrifice experimental control but are more ecologically valid, as people are asked to remember events from their own lives as opposed to lab-created stimuli. With the differences between the two types of memory, many studies have been done to identify and comparatively analyze the neuroanatomical correlates between the two approaches. In fact, the literature on LM and AM mostly shows there to be very little to no overlapping between the activated regions (McDermott et al., 2009).
Two important factors contributing to the differences in neural correlates are the varying levels of the involvement and impact of emotion, vividness and temporal order on AM and LM. Unlike LM, AM tends to have vivid sensory details and rich emotional context due the real-world environment and context in which they are formulated. It has been found that emotion contributes to AM retrieval (amygdala activation), even before event-specific memories are completely formed, whereas vividness develops late (occipital and precuneus activation) as reflexive processes turn to recovered visual images (LaBar, K.S. et al., 2005). Meanwhile, a study on temporal-order memory for autobiographical events has found AM to involve both reconstruction and distance processes (more activation in the left dorsolateral PFC), and LM to primarily involve distance processes (more activation in the right dorsolateral PFC) (Jacques, Peggy St. et al., 2008).

The Study

The particular study I have been involved with this quarter sought to explore the memory retrieval and neural correlates of AM across various conditions, such as the source of the event sequences (first-person memory or third-person perspective), whether the event sequences have been pre-exposed or not, and the temporal order of the event sequences. More specifically, this was done using naturalistic stimuli derived from participants’ own lives through the use of automatic wearable cameras—a method that has been used by very few AM experiments in slightly different contexts (F. Milton, 2011). Through the use of these devices, the procedure is able to reduce the intentional encoding of event sequences to be tested on, and capture the varying levels of contextual details between images—allowing for stronger insight into everyday recognition memory.
In order to generate enough events for naturalistic stimuli for the experiment, 18 young participants wore an unobtrusive wearable digital camera for 3 weeks that automatically took photographs whenever its sensors detected change in the environment. For each participant, 120 event sequences were captured over the course of 3 weeks (40 per week)—with an event sequence consisting of 8 photographs of an event taken within 15 minutes. One week later, each participant went to the lab to preview 60 sequences from their own life as well as 60 from another participant’s life (the events’ sources were unknown to them), for a total of 120 event sequences. Upon previewing the images, each participant encoded the event sequences by rating them based on their perceived distinctiveness. One day later, each participant viewed 240 event sequences (120 from their own life and 120 from another participant’s life) while undergoing fMRI scanning.

There were eight experimental conditions with 30 event sequences each—these were a combination of the following “variables”: ‘Source’ (Self vs. Other), ‘Preview’ (Previewed vs. Non-Previewed), and ‘Intactness’ (Scrambled vs. Intact). For the scrambled sequences, the temporal order of the last four photographs was rearranged while the first four were kept intact. By scrambling the last four photographs, the neural correlates of associative novelty (familiar objects or stimuli arranged in unfamiliar configurations) could be investigated. For each of the 240 trials, each participant had to chose one of four responses on whether each event sequence was derived from their own or another participant’s memory, and on whether each event sequence was temporally altered or not.

Through multi-voxel pattern analysis (MVPA) and the utilization of networks of interests from McDermott et al.’s 2009 meta-analysis—the autobiographical network (associated with retrieval of AM) and the retrieval success network (associated with LM recognition), the study
found that: (1) The neural signatures of memory for personally-experienced events were distinct from recognition of event photos seen in the lab; (2) The autobiographical network excelled in decoding memory source (self vs. other) while the retrieval success network was better at decoding pre-exposure status; (3) Ultimately, the networks have different specializations with regards to recollection-based memory and familiarity-based memory.

*My Role: Behavioral Data Coding and Analysis*

When I joined the lab early this quarter, the experiment itself had already been completed, along with some fMRI data analyses that had produced some insightful findings. However, there had not yet been much behavioral data analyses, as the initial focus was on the fMRI data and the analysis of the neural correlates of AM. I spent most of this quarter coding scripts and functions in MATLAB that could dynamically pull, format, organize, manipulate and analyze critical raw behavioral data collected from the experiment. Here is a breakdown of some of the important functionalities the scripts provide so far: (1) The ability to pull organized Preview data, Scan data or both data sets of a subject (based on the subject input) in an array format (converted from .mat) and save the data; (2) Calculation of the Scan hit rates (participant’s response accuracy rate) for any given subject across all eight conditions; (3) Calculation of the Scan hit rates for any given subject for only Self vs. Others and for only Intact vs. Scrambled; (4) Calculation of the low and high distinctiveness rating percentages for any given subject for event sequences that were previewed. Other functionalities on the roadmap include the calculation of the Scan hit rates for each participant per week (Self vs. Other and Intact vs. Scrambled), the pulling of data from all of the subjects at once, and some more dynamic formatting and organization of the data.
Once all of this is finished, we will be using SPSS software to start analyzing and drawing insights from the refined and organized data, and hopefully draw ties to the prior fMRI findings as well. A particularly interesting set of data is the distinctiveness ratings from the Preview trials—for example, it would be interesting to see if there are any correlations between the low and high distinctiveness ratings of the Preview trials and all of the other conditions, or perhaps if there are neural correlates for the distinctiveness ratings with regards to the autobiographical and retrieval success networks as well. It would also be interesting to see whether event sequences derived from a participant’s own memory tend to have a higher distinctiveness ratings than those derived from another participant’s memory—as one would assume based on intuition.

References


Retrieval tasks differ substantially in their neural substrates.” *Neuropsychologia.*

47, 2290-2298.


733–744.


Effects of Form of Stimulus on Analogical Problem Solving

Yao (Sherry) Hsiao

**Introduction**

Analogical reasoning allows human to form generalizations between similar systems, recognize underlying abstract concepts, generate schema, and adapt the knowledge to the novel yet analogous situations. This is evident in their study researched by Schunn (1999), he found that priming was the key factor that allow people to relate two different systems, it was effective even without the awareness of the person. The related systems that we have more knowledge about, is called the source system; while the unknown system that is seemingly novel but analogous to the source system is called target system or problem. As one spontaneously retrieves the source and applies the knowledge to a target problem, the ability is called spontaneous transfer. When one is hinted with the relevance of the source system to the target problem, the ability to draw inferences is called total transfer. From the previous research done by Gick and Hoyak (1980), we see that subjects were better at solving the target problem with a source analog than without, and the transfer rate was higher when a hint was given. Another prior study, investigated by Dunker (1945), had probed into the transfer phase of forming an analogy between systems through the famous radiation problem. The radiation problem is a target problem in which a doctor has to treat a patient’s tumor with a type of radiation instead of surgery. If the intensity of the radiation is high, the tumor can be destroyed along with the healthy tissues around; if the intensity is low, the radiation will not remove the tumor or harm the healthy tissues. Only ten percent of the subjects were able to solve the problem without a source analog. Most research in analogy centralizes the discussions around the mapping and inference
stages in the reasoning process. In this experiment, we were most interested in finding what source presentation characteristics are most important and effective in facility spontaneous transfer from a source system to a novel target problem, utilizing the radiation problem as the target problem as well.

**Method**

**Participants**

A total of 126 subjects were enrolled in the study. All the subjects were the undergraduate students in University of California, Los Angeles who enrolled in psychology courses that required students to experience being a participant of an actual ongoing experiment. They were recruited from the UCLA Psychology Department Subject Pool via SONA system and were granted 1 credit for the course they were enrolled in after they had completed the experiment. Subjects were randomly divided into three conditions including verbal, verbal diagram, and verbal animation. Participants were blind to the objective of the study. Those who had pre-exposed to the elements of the experiment (students who had taken the courses of Psychology 120A or Psychology 85 had learned about the radiation problem) were declined to participate due to ineligibility. The experiment was conducted during the Fall quarter of 2014 academic year.

**Materials**

The experiment was primarily computer –based. Computers with ViewSonic CRT monitors of resolution 1024 x 768 pixels were utilized in the research lab. Headphones or speakers were also used when spoken monologues were administered during the experiment. Subjects engaged in the experiment from an approximately 65-centimeter viewing distant. The
entire experiment was presented and collected using an online survey system called Qualtrics that was designed for research and experiments. All the animations were created using Psychtoolbox 3 or MATLAB and inserted into the survey. Questions administered were presented in multiple choice or free response form.

Design

Subjects were divided into verbal, verbal diagram, or verbal animation conditions and were presented with the same four explanatory scenarios in ways according to their conditions. A number of 42 subjects had participated in each condition. In the verbal condition, subjects were given the spoken monologue only, explaining what happens in each scenario. Subjects had to listen to the auditory explanations closely in order to comprehend the scenarios. In the verbal diagram condition, subjects were also given the spoken monologue but this time with an explanatory diagram. The pictorial supplements for the four scenarios were presented as visual aid to the auditory descriptions. These pictures were colored but static. In the animation condition, subjects received animations in addition to the auditory-verbal account. The supplemental animation of each scenario was incorporated with the same spoken monologue explaining the scenarios. The animations were the moving version of the static diagrams presented in the verbal diagram condition. The three conditions were the different methods of presenting the same information about the four scenarios.

The four source scenarios had same objective. These scenarios depicted one or multiple cannons firing cannonballs at the enemy octagon inside a friendly barrier. The goal of the task was to defeat the enemy octagon effectively with maximum damage but minimum harm to the friendly barrier around it. In order to completely destroy the enemy octagon without destroying
the friendly barrier, the amount of cannons (one or more cannons) and the size of cannonballs (small or large) were different among the scenarios as different approaches to achieve the optimal result. In scenario 1, a single cannon fires small cannonballs at the enemy octagon, the small cannonballs passes through the friendly barrier harmlessly and cause no noticeable damage to the enemy octagon. In scenario 2, a single cannon fires large cannonballs at the enemy octagon, subjects learn that cannons can shoot large cannonballs. One fire at a time, the large cannonballs inflict minor damage to the enemy octagon but major damage to the friendly barrier, the large cannonballs break through the friendly barrier before completely destroy the enemy octagon. In scenario 3, subjects learn that more than one cannon can be used to fire at the enemy, multiple cannons surround the enemy octagon and fire large cannonballs at the enemy octagon from multiple directions. The large cannonballs pass through the friendly barrier, exerting major damage to it at multiple sites; the cannonballs then converge at the enemy octagon simultaneously, also inflicting major damage to it. The large cannon balls in this scenario too break through the friendly barrier before the enemy octagon is fully wrecked. In scenario 4, multiple cannons fire small cannonballs at the enemy octagon. In this scenario, the cannons are also positioned around the enemy octagon, and cannonballs pass through the friendly barrier converging at the enemy octagon. The small cannonballs cause harmless damage to the friendly barrier individually at different sites but moderate damage to the enemy octagon when they converge at the same time. Subjects learn that converging small cannonballs are able to damage the enemy octagon after observing this scenario. Among the four scenarios, only scenario 4 is capable of successfully and effectively destroying the enemy octagon without harming the friendly barrier.
After the subjects observed the scenarios according to their conditions, scenario assessment was administered. A free response question of why the single or multiple cannons failed or succeeded destroying the enemy octagon while maintaining minimal damage to the friendly barrier was asked. Followed by two multiple-choice questions, the level of damage to the enemy octagon and friendly barrier were asked. Each question was asked per scenario thus each subject should answer a total of 12 questions.

A shortened version of Raven’s Progressive Matrices test consisting of 12 questions was utilized as the filler task. Subjects were asked to identify the missing element that completes a pattern as a whole. The Raven’s test not only served as a filler task that distracts subjects from developing associations between the source scenarios and the radiation problems, it also measured the general intelligence of each subject. The score was utilized to correlate the general intelligence to the solving rate of the radiation problem later in the study.

Subjects then were asked to solve the radiation problem, first with no hint to the relevance of the source scenarios in solving the problem, then same question was asked again with a hint. It first assessed the subjects’ ability to spontaneously transfer knowledge from the source scenarios to the target problem without a hint. Then assessed the subjects’ ability to directly transfer knowledge from the source scenarios to the target problem after a hint recalling the four scenarios. A survey of questioning the subjects whether they had resembled the radiation problem with the four scenarios before the hint was given. The survey also asked if the subject had seen the radiation problem somewhere else before they participated in our experiment.

Three research assistants graded the subjects’ answers according to a series of scoring guidelines. Two research assistants first graded the responses and a third research assistant
graded again on the responses where the scores did not agree. The scores had reflected subjects’ understanding of the four source scenarios, and whether they had solved the radiation problem correctly with or without a hint.

**Procedure**

Participants arrive at the lab according to the time they signed up on the SONA system. After confirming participants’ name and made sure they have never taken the courses that potentially expose the objective of this experiment, research assistant led participant to a room, explained the procedures and began the experiment. Participants were first introduced the four scenarios according to their conditions, followed by the scenario assessment, Raven’s Progressive Matrices test, radiation problem without hint, radiation problem with a hint, and finished with a survey. The time limit to complete the experiment was one hour; many nonetheless finished around 30 to 40 minutes. Participants were granted one course credit upon their completion of the experiment.

**Results**

The radiation problem was evaluated according to their comprehension of the convergence concept. If subjects demonstrated two out of the three criteria, including multiple radiation sources, low intensity of radiation sources, and convergence, they were considered correct. Their responses were graded both before and after the hint and the scores were used to evaluate subjects’ ability to spontaneously or directly transfer the knowledge from source scenarios to the radiation problem. A Pearson chi-square test was utilized to compare the radiation problem transfer rate across the conditions. The spontaneous radiation problem transfer rate in the Verbal Animation condition was significantly higher than the transfer rate in the
Verbal Diagram condition \( \chi^2(1, N = 84) = 10.50, p = .001 \) and Verbal conditions \( \chi^2(1, N = 84) = 8.02, p = .005 \); however it was not significantly different between the Verbal and Verbal Diagram conditions \( \chi^2(1, N = 84) < 1 \). The total radiation problem transfer rate in the Verbal Animation condition was significantly higher than the rate in Verbal Diagram condition \( \chi^2(1, N = 84) = 4.53, p = .03 \), and was not significantly different from the Verbal condition \( \chi^2(1, N = 84) < 1 \). The total radiation problem transfer rate was also not significantly different between the Verbal and Verbal Diagram condition \( \chi^2(1, N = 84) < 1 \). The results suggest that providing animated representations rather than static pictorial representation increased the subjects’ spontaneous retrieval of the relevant source analogs. Although the comparison was not significantly different, the Verbal Diagram condition total transfer rate was slightly lower than the Verbal condition, suggesting that pictorial representations were not as effective as animated representations both before and after the hint.

An independent sample t-test was used to compare the free response and multiple-choice questions. Significant effects of conditions were found when comparing the free response scores. Subjects in Verbal Diagram condition scoring higher than subjects in Verbal condition \([t(82) = 2.23, p = .03]\), and Verbal Animation condition subjects scoring higher than Verbal Diagram condition \([t(82) = 2.27, p = .03]\) as well as Verbal condition subjects \([t(82) = 27.00, p < .001]\). A significant effect of condition was found when assessing the multiple-choice scores. Verbal Animation subjects scored higher than Verbal Diagram subjects \([t(82) = 2.76, p = .007]\) and Verbal subjects \([t(82) = 3.07, p = .003]\). This result suggested that participants learn more and transfer better when the scenarios were presented in a verbal-animation incorporated form.

Pearson chi-square statistic test was utilized to compare the spontaneous transfer rate to the radiation problems in relation to the Raven’s Progressive Matrices test scores. Subjects either
scored low (< 8 correct) or high (> 8 correct) on the Raven’s test. Among the low score subjects, the transfer rate in the Verbal Animation condition was significantly higher than in the Verbal Diagram [$\chi^2(1, N = 30) = 7.75, p = .005$] and Verbal [$\chi^2(1, N = 33) = 6.64, p = .01$] conditions. No difference between the Verbal Diagram and Verbal conditions [$\chi^2(1, N = 27) < 1$] was observed. Among the high score subjects, transfer rate did not differ across condition [$\chi^2(1, N = 38) = 3.64, \chi^2(1, N = 38) = 1.58, \chi^2(1, N = 38) < 1, p > .05$]. These results suggested that subjects who scored lower on the Raven’s test were benefited more from the supplemental animations than those who scored higher.

**Discussion**

Analogous learning is an effective method of learning by allowing learners to have a profound and conceptual comprehension of a source domain and transfer the knowledge to a less known one. This experiment explored which representations of the source system affect subjects’ understanding of the system itself and the rate of transfer to a target problem the most. The results revealed that supplemental animated representation with spoken monologue, when comparing with static pictorial representation, best enhanced learning of the source and transfer rate to the isomorphic problem. One limitation of this study could be attributed to the nature of grading by research assistants. Grading the subjects’ responses was a judgmental task although a third grader was employed to eliminate the discrepancy. Different opinions of the scoring could yield in different experimental findings. A simple solution to this problem is to change the free responses questions to multiple-choice form. Based on the results found in this study, further research should investigate on whether animated stimuli is able to facilitate subject’s true understanding of the system concept underlying the target problem. Also, further research may also want to investigate whether animation without spoken monologue can still generate the
same effect we found in this study. Any further investigations on the topic of analogy will contribute to the knowledge of how we learn, as humans, through analogous reasoning.

References

