

Spring 2014

Table of Contents

<i>Flash-Lag Effect with Biological Walker Motion</i>	3-9
Andy Huang, Faculty Mentor: Hongjing Lu, Ph.D.	
<i>Prospective Representations of the Future</i>	10-15
Brittney Medina, Faculty Mentor: Jesse Rissman, Ph.D.	
<i>When Should We Correct? : Feedback Timing and the Error Generation Effect</i>	16-22
Charlie Hall, Faculty Mentor: Alan Castel, Ph.D.	
<i>Active vs. Passive Learning</i>	23-28
Emma Murray, Faculty Mentor: Phil Kellman, Ph.D.	
<i>Learning and Perception of Music</i>	29-35
Erik Shiboski, Faculty Mentor: Phil Kellman, Ph.D.	
<i>Analogical Reasoning with Rational Numbers</i>	36-44
Eugene Goh, Faculty Mentor: Patricia Cheng, Ph.D.	
<i>Effects of Active and Passive Learning on Cognitive Load</i>	45-53
Giancarlo Sanguinetti, Faculty Mentor: Phil Kellman, Ph.D.	
<i>Adaptive and Fixed Schedules in Complex Fact Learning</i>	54-66
Jason Ku, Faculty Mentor: Phil Kellman, Ph.D.	
<i>Remember-Know Judgments and the Attention to Memory Hypothesis</i>	67-72
Kaitlin Swinnerton, Faculty Mentor: Jesse Rissman, Ph.D.	
<i>Foreign Language Learning Through a Virtual Environment</i>	73-80
Katherine Guardardo, Faculty Mentor: Jesse Rissman, Ph.D.	
<i>Videogames & Motor Learning</i>	81-86
Kylie Springsteen, Faculty Mentor: Zili Liu, Ph.D.	
<i>Effects of Active and Passive Presentation Style on Factual Learning</i>	97-93
Lorena Hinckley, Faculty Mentor: Phil Kellman, Ph.D.	

<i>Confidence Weighting on Multiple Choice Tests</i>	94-100
Monica Vu, Faculty Mentor: Elizabeth Bjork, Ph.D.	
<i>Causal Parsimony and the Appreciation of Fractal Artwork</i>	101-115
Priyanka Mehta, Faculty Mentor: Patricia Cheng, Ph.D.	
<i>Virtual Reality and Cognitive Mapping Strategies in Language Acquisition</i>	116-122
Taylor Henry, Faculty Mentor: Jesse Rissman, Ph.D.	
<i>Differences in Learning across Criterion Trials in Adaptive Sequencing and Fixed Expanding Schedules</i>	123-134
Victoria Groysberg, Faculty Mentor: Phil Kellman, Ph.D.	

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

Flash-Lag Effect with Biological Walker Motion

Chin-Wei Andy Huang

The flash-lag effect is a visual effect that occurs when a stimulus is flashed in alignment with a continuously moving object resulting in the flash to appear to be lagging behind the moving object. This effect has been widely studied with the purpose of understanding why this effect occurs and to what purpose.

In the review work by Bart Krekelberg and Markus Lappe (2001), both researchers analyze the different hypothetical explanations for this phenomenon. Early researchers believed that the flash-lag effect results from the brain perceiving the flash and the motion of the object, but because of neuronal lag, has to extrapolate the position of the moving object. However, this theory is insufficient because it cannot explain the lack of predicted overshoot that would occur if the moving object were to suddenly shift position. Another explanation given is that the visual system responds with shorter latency to moving stimuli rather than to flashed stimuli. In this way, the differential latencies hypothesis states that the phenomenon is a by-product of how the visual system is setup, but the evidence is lacking that this hypothesis can explain fully the different flash-lag effects across all its different dimensions. Another school of thought, which begins to treat the actual nature of flash as special, is the temporal integration, which proposes that the visual system collects position signals over time and estimates the position. The flash is important because the information from the stimulus persists for a time and biases the position estimate towards that last scene position. Where this also diverges from the motion extrapolation hypothesis is that the temporal integration model is able to account for reversing or slowing down of the moving object and its effect on the flash lag. Finally, Krekelberg and Lappe review the Postdiction model, which explains that the brain constructs a percept by combining the

internal model of the world with current external input. The flash, in this way, is a high salience external input that resets the internal model of the world. After this “reset”, the internal model must be rebuilt again and because the moving object has moved on, the position of the object is thought to be past the position of the flash. (Krekelberg and Lappe, 2001)

Continuing in this line of research, Nijhawan (2002) explores similar explanations, but also discusses the possible biological reasons to have the primate visual system set up in this manner. There are a number of interesting effects that occur based around this flash-lag effect, which include flash-terminated and flash-initiated cycles. Flash-terminated cycle is defined as when the pre-flash trajectory of the moving object matches a regular flash lag display, but when the flash disappears so does the moving object. In this case, there is no flash lag effect. In the flash-initiated cycles, the motion of the object is started by the flash, but in this case there is a flash lag effect. In his work, Nijhawan (2002) explains that the reason for our visual system’s accurate representation of position at motion-termination was evolutionarily important in how people were able to hunt and find food or resources. On the other hand, this process, also known as backward masking, of identifying flash-terminated cycles accurately seems to allow for the visual system to have some leeway in developing the flash-lag effect. (Nijhawan, 2002).

In Katsumi Watanabe’s and Kenji Yokoi’s work (2006), the researchers were able to find that the mislocalization of the flash was not symmetric around the moving stimulus, but rather anisotropically mislocalized. This meant that the perception of the amount of the lag of the flashing stimulus was not equal around the moving object, but rather flashed stimulus that appeared farther ahead of the moving object experienced greater flash-lag effect. Furthermore, flashed stimulus that appeared behind the moving object did not experience an equal lag effect. In fact, in their study, Watanabe and Yokoi (2006) found that the perceived flash-lag effect

seemed to converge onto a singular point somewhere behind the moving object. In their second experiment, Watanabe and Yokoi explored whether or not such an effect was caused by a perceived warping of the space around the object in motion. However, as the result in the experiment showed, there was no perceived warping in the space, but there was a similar effect in that it seemed the flash-lag effect was converging onto a singular point (Watanabe and Yokoi, 2006).

All in all, our lab decided to focus its experiments on the flash-lag effect especially in how it deals with biological motion. As stated above, the foundation of our visual system seems to be built upon biological reasons of being able to better find resources, which may mean that biological motions, like walking or running, is processed differently than normal motion. In this study, we hypothesized that biological motion is processed differently when it comes to the flash lag effect.

Method

Participants

There were 31 participants. They were all students from the University of California, Los Angeles that had signed up for the experiment using the SONA system. All participants received course credit for participating in the experiment.

Design and Procedure

All the participants were given a visual experiment task where they had to make judgments on the perceived misalignment of the flash relative to certain limbs of a walking figure that was presented on the screen. Prior to starting the experiment, the participants were asked to sit with their head about 57cm in front of the monitor they would be completing the task on. During the task, the participant would see a walking figure that was a half-green and half-

blue and a dot would flash on the screen. The participant would either hit the left arrow key or the right arrow key dependent on which side, relative to the blue limb, the dot appeared.

The dot could appear at seven different locations around the blue hand or at seven different locations around the blue leg. The dot would always appear at the tip of the walker's blue limb on frame 53 of the walker in the walking cycle. These positions were time offsets of the tip of the limbs in seconds. For example, at the 2 offset for the hand, the dot would flash at where the hand would be 2 seconds later. The experiment randomized around which limb the dot would appear, but each participant saw an equal number of dots appear either around the hand or the foot of the walker.

In the experiment, there were 2 blocks of trials that the participant would go through. Each block had 140 trials for a total of 280 trials in the experiment. In the first block, the participant would either see the walker walking backwards or forward, but always the same way for the entire block. The direction of the walker would flip going into the second block. Participants were randomly placed into which condition (forward or backward walking walker) they saw first. Before the actual experiment, there were 10 practice trials that had trials exactly like the ones in the testing phase, but these trials ran about two times slower during the practice trial.

At the very end, all participants were given a questionnaire to find the participant's autistic quotient score. This was key because there has been previous work done suggesting that autism can affect the ways that people perceive biological motions (van Boxtel and Lu, 2013).

Results and Discussion

As seen in Figure 1 (below), each line is shifted toward the positive time offset, which is evidence of the flash lag effect because it demonstrates that the participants on average were

saying that the flashed dot appeared on the opposite side of the limb than it actually did appear. The results demonstrate that participants demonstrated a larger flash-lag effect for the foot going forward than when the foot was going backward. On the other hand, there doesn't seem to be the same significant difference in effect when it comes to the hands whether going forward or backward. Such results may suggest that there are special mechanisms within the visual system that deal with the motion of legs, but why it has such a larger effect on forward moving feet is unknown.

With these results, future studies will focus on explaining why there is such a large effect on the forward moving feet rather than when the backward moving feet and why this is not seen in the hand movements. Furthermore, future studies will include studying the effects of speed on the flash lag effect as limbs do not always move at a constant speed naturally.

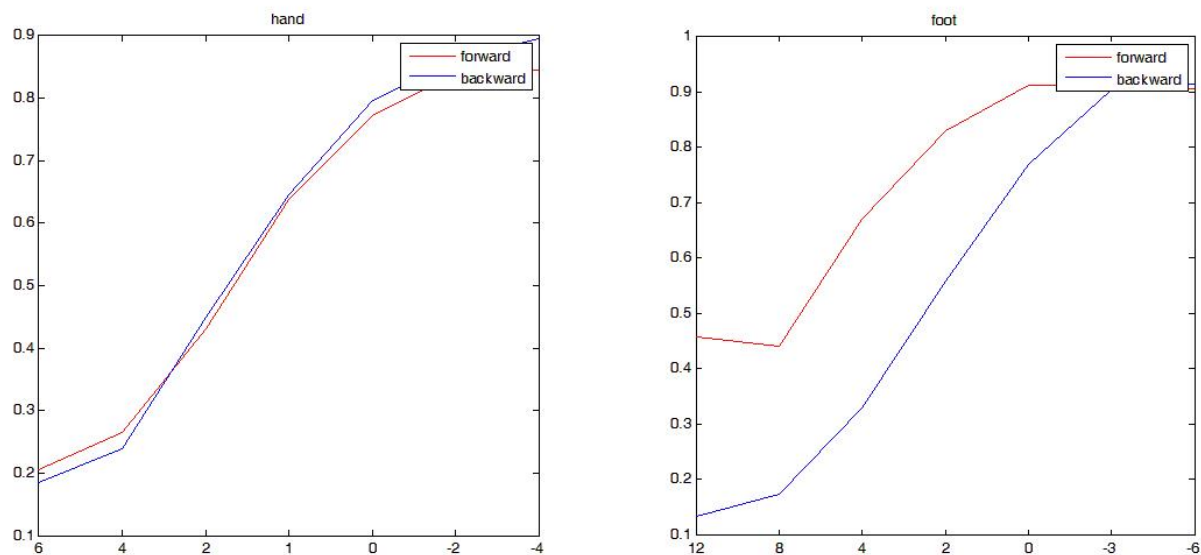


Figure 1. Plots of average proportion of responses given that corresponded to the direction of the motion of the walker. One can see that there is a significant increase in responses for the

forward-facing walker's foot at all positive time offsets as compared to the responses for the backward-facing walker's foot.

References

Krekelberg, B. and Lappe, M. (2001). Neuronal latencies and the position of moving objects.

TRENDS in Neurosciences, 24(6). 335- 339.

Nijhawan Romi (2002). Neural delays, visual motion and the flash-lag effect. *TRENDS in*

Cognitive Sciences, 6(9). 387- 393.

van Boxtel, J. J. A and Lu, H. (2013). Impaired global, and compensatory local, biological

motion processing in people with high levels of autistic traits. *Frontiers in Psychology*, 4.

1- 10. doi: 10.3389/fpsyg.2013.00209

Watanabe, K. and Yokoi, K. (2006). Object-Based Anisotropies in the Flash-Lag Effect.

Psychological Science, 17. 728-735. doi: 10.1111/j.1467-9280.2006.01773.x

Prospective Representations of the Future

Brittney Medina

Memory is designed such that the simple act of recalling a past event changes your memory of that event. Each time an event is remembered, past events are reconstructed in the mind differently than how it was encoded in the first place. Why is the mind designed this way? Why are our memories never set in stone, so to speak? Many studies have shown that the same neural mechanisms for remembering the past are also involved when trying to imagine the future (Schacter et al., 2007). For this reason, it is important to understand the idea of prospection and the ways in which the mind creates representations of our possible futures. I will discuss a few paradigms that explore this phenomenon as well as briefly introduce a proposed project in the Rissman Memory lab that may be able to delve into the neural mechanisms of prospective cognition.

What is prospection? Prospection is the ability of the human brain to imagine non-existent events and “pre-experience” possible future events that may happen (Mullally & Maguire, 2013). This includes planning, predicting, hypothetical scenarios, teleological patterns or daydreaming about future events. It has been argued that episodic memory supports the remembering of past events as well as the ability to engage in “mental time travel” (prospection) (Schacter et al., 2007). The leading question in prospection research, then, will explore the ways in which past experience is used in forming future representations, and what other factors might influence them.

Episodic memory, the memory for our everyday personal experiences, is widely known as a constructive process rather than a reproductive process in that it is prone to errors and

illusions. This fact, however, is an important aspect because it allows for a system that can draw upon past events to piece together plausible future episodes in a manner that is a flexible recombination of the past. This is more commonly known as the *constructive episodic simulation hypothesis* (Schacter & Addis, 2007).

Neuroimaging studies have shown that common brain systems are involved in remembering the past and imagining the future and are evidence for the constructive episodic simulation hypothesis: In one study, participants were instructed to talk freely about the near or distant future while a positron emission tomography scan was carried out. The scan showed overlapping activity during descriptions of the past and future events. These overlapping regions included the prefrontal cortex, hippocampus and parahippocampal regions (Schacter et al., 2007). A different study made use of fMRI techniques in which participants were instructed to remember specific past events, imagine future events or imagine specific events that involved a familiar individual in response to event cues. This study was interesting because the remembrance of specific past events and the imagining of future events showed neural activity overlap; however, these same regions were not activated when imagining events involving Bill Clinton. This shows that a specific neural mechanism is responsible for the construction of events in one's personal past or future. Lastly, another fMRI study consisted of asking participants to generate a past or future event cued by a noun during a construction phase. Once the participants had the event in mind, they elaborated on the event for 20s. Analysis showed that the left hippocampus was engaged by past and future event construction (Addis et al., 2006). Integrating the data found from these three studies of autobiographical memory, strong evidence suggests that remembering the past and imagining the future is associated with a specific core brain system that integrates information from past events in order to construct the mental

representations of future events. This core brain system allows one to shift from perceiving the immediate environment to an alternative perspective.

More specifically, the core brain system involves four core kinds of mental simulations that are closely linked to the *default mode network*. The default mode network is a network of brain regions that are active when the individual is not focused on external stimuli and relies heavily on introspection. These four mental simulations involved with the default mode network are: navigational, social, intellectual and memorial (Seligman et al., 2013). The first mental simulation occurs when one tries to spatially navigate through an environment: Imagine walking from your house to campus. The second is social simulation: Imagine a conversation with a friend about going to the movies and your response to the conversation. The third is an intellectual simulation: Imagine yourself critically thinking about an article and evaluating its methods or deciding if the paradigm has external validity. The fourth is memorial: recall an event that went very badly or unexpected and imagine a different outcome to this scenario (Seligman et al., 2013). Each of these scenarios demonstrates one's capacity to generate and explore endless situations and alternatives to the present. A study recently suggested that these mental simulations all have one common element: they all involve a kind of "*scene construction*" or "*scene building*" (Hassabis & Maguire, 2007).

This scene construction is defined as the process of mentally generating a coherent scene through the retrieval and integration of the relevant informational components. It is argued that much of the overlap seen in neuroimaging studies is due to this core brain system that actively engages in scene construction of events (Hassabis & Maguire, 2007). This scene construction theory places more emphasis on the role of the hippocampus in imagination, spatial navigation and memory for past events whereas the constructive episodic brain hypothesis has a broader

focus on the core network. Scene construction theory, however, does not suggest that the hippocampus is solely responsible for episodic memory but that it does play a crucial role human's ability to engage in prospection (Mullally & Maguire, 2013).

With these hypotheses in mind, future prospective research would benefit from exploring the questions of how we learn and remember our past experiences to make informed predictions about the future and the role of the hippocampus in these processes. A proposed project in the Rissman Memory will try to delve into two sources of prospective insight: episodic and schematic sources. For instance, when drawing upon an episodic memory, hippocampal regions may trigger engagement of medial PFC in order to project oneself into the emerging representation (St Jacques et al., 2011). The second source, drawing upon semantic/schematic knowledge, engagement of the lateral PFC may interact with the hippocampus to contribute to the construction of a scene-like 'guess' for future events. These sources will be studied through the use of wearable necklace-mounted *Autographer* cameras for participants to wear daily during a 3-week period. During this period, the Autographer will take thousands of pictures of the participant's life, which will be compiled into individual sequences of events. Later, while in the scanner, participants will view rapidly presented sequences of photos from their own life or sequences from someone else's life. After viewing the 6 initial photos of an event, the screen will go blank for 4s and participants will be instructed to construct a mental representation of what might come next. Afterwards, 6 more photos will be presented that either depict the continuation of that event in a temporally intact fashion, the continuation of that event in a temporally scrambled fashion or images from a completely different event. Participants will, then, have to make judgments about whether the second half of the event was intact, scrambled or new. This paradigm is unique as it involves repeated sampling of subjects' experiences in real time and in

subjects' natural environments. Pilot data has been collected from 5 participants performing a variation of this task and shows promising results. In essence, this project has the potential for elucidating the neural mechanisms responsible for prospective cognition.

In conclusion, as well as understanding the neural mechanisms driving prospection, it is important to understand the applications of the constructive processes that go beyond just planning for and anticipating possible future events based on past experience. These processes also aid in the imaginative and creative abilities needed for general problem solving. The next question, then, will be: Can we teach the next generation of young people to be better simulators, evaluators, and implementers of the future?

References

- Addis, D., Wong, A., & Schacter, D. (2007). Remembering The Past And Imagining The Future: Common And Distinct Neural Substrates During Event Construction And Elaboration. *Neuropsychologia*, 45(7), 1363-1377.
- Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in Cognitive Sciences*, 11(7), 299-306.
- Mullally, S. L., & Maguire, E. A. (2013). Memory, Imagination, and Predicting the Future A Common Brain Mechanism?. *Neuroscientist*, 20(3), 220-234.
- Schacter, D. L., & Addis, D. R. (2007). The Cognitive Neuroscience Of Constructive Memory: Remembering The Past And Imagining The Future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1481), 773-786.
- Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: the prospective brain. *Nature Reviews Neuroscience*, 8(9), 657-661.
- Seligman, M. E., Railton, P., Baumeister, R. F., & Sripada, C. (2013). Navigating Into the Future or Driven by the Past. *Perspectives on Psychological Science*, 8(2), 119- 141.
- St. Jacques, P. L., et al. (2011). Watching my mind unfold versus yours: an fMRI study using a novel camera technology to examine neural differences in self-projection of self versus other perspectives. *J Cogn Neurosci*, 23(6), 1275-1284

When Should We Correct? : Feedback Timing and the Error Generation Effect

Charles D. Hall

Introduction

Learning can occur in vastly different ways: observation, mimicry, reading, memorizing, and even testing. Testing, in addition to being an effective means of evaluating learning, also serves as a learning event itself. Tests have been found to significantly increase learning by activating the associative networks that prevent forgetting, with or without feedback. There is little argument that testing events contribute to effective long-term retention more than simple exposure to the information. However, the degree of successful retrieval for particular items in a testing event may vary considerably. Successfully retrieving items from memory is certainly beneficial for long term retention, but researchers do not agree on the effects of unsuccessful or incorrect retrieval and its benefit on the long-term retention for the correct response. Of further interest, the timing and degree of feedback after testing events have been known to affect retention, especially for incorrect responses.

The Error Generation Effect

When testing occurs, the act of generating or retrieving a response is known to strengthen the associative network of the cue and increase the probability of recalling that response again at a later time. Given that information, researchers have attempted to find out how the generation of false responses might affect learning. Researchers who hypothesized that error generation may increase retention for the correct cue rely on the notion that any act of generation, regardless of correctness, will activate the associative network of the cue and prepare the learner to remember the correct cue better once feedback is given. On the other hand, researchers who believe that error generation may be detrimental to the retention of the correct response rely on the idea that incorrect generation creates stronger and more significant interferences with incorrect associations than simply studying the correct answer before a final testing event. In order to test the effects of error generation, 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.

Although this study and its commentary on the effects of generating errors in testing events were significant, other research has challenged and modified its claims, specifically in relation to the timing of feedback given for incorrect guessing. Vaughn & Rawson produced results suggesting that incorrect guessing led to worse performance and that the number of incorrect guesses did not alter this effect. They also manipulated when restudy occurred after an initial incorrect guessing test. In their first experiment, Vaughn & Rawson used the same methods as Kornell et al, except that they manipulated the number of pretrial incorrect guesses or pre-studies. The results suggested that word pairs studied read-only were better remembered than word pairs in which participants generated incorrect guesses. The next experiment reveals the important role feedback plays in moderating the effects on incorrect guessing. The experiment was designed similarly to Experiment 1, with the addition of an immediate study group. This group received a particular word pair's study phase directly after the pretrial guess or pre-study. In other words, if a participant was asked to provide an incorrect guess in a pretrial guess condition, it would be immediately followed by the correct word pairing. In this experiment, delayed study of word pairs resulted in worse performance for guessing than pre-study material; however, if the study session occurred directly after a pre-trial guess or pre-study, than guessing resulted in greater final performance. This suggests that the timing of corrective feedback may play a critical role in the effectiveness of forced guessing. Furthermore, the results indicated that the pre-study conditions benefitted from traditional spacing effects, about fifteen percent, whereas guessed items actually decreased in performance with time. In their third experiment, Vaughn & Rawson attempted to examine an explanation of the delayed and immediate study interaction experienced by guessed answers.

They postulated that for prestudy (read-only) conditions, immediate study resulted in a single massed practice session and delayed study resulted in two separated spaced practice events. Similarly for guessing, immediate study would result in one practice session in which the guess acted as an intermediate step between the cue and target while the delayed study condition acted as two study conditions: one for the cue-guess and one for the cue-target. If this was true, the ineffectiveness of guessing trials could be attributed to source monitoring issues where participants cannot differentiate between their guess and the correct answer. In order to examine if source monitoring could play a role in the error guessing phenomena, Vaughn & Rawson conducted a similar experiment to Experiment 2 while adding an additional question in the final test phase that asked participants to indicate if they remembered their incorrect guess. The results were identical to Experiment 2 for final performance. Interestingly, participants were extremely accurate in providing their guesses before the correct answer, which indicates that they did know the source of their responses. This finding suggests that when participants guess with immediate feedback, something beneficial is occurring beyond just preventing confusion between their guess and the correct answer. It is likely that the pre-trial guess activates the associative network of the cue which allows the correct answer to be encoded more deeply than when simply studied in a read-only manner. This implies that, with proper feedback, asking learners to generate guesses will benefit them if and only if their incorrect guesses are immediately corrected by feedback

References

- Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 989–998.
- Vaughn, K.E. & Rawson K.A. (2009) When is guessing incorrectly better than studying for enhancing memory? *Psychon Bull Rev* (2012) 19:899–905 DOI 10.3758/s13423-012-0276-0

My Experience

I had a great time in the lab this quarter and was pleased that I could take on extra work and take advantage of extra opportunities available to me. I was able to participate in Science Poster Day, which challenged me to write an academic abstract, design a poster, and engage in discussions about research. Carole was incredibly helpful, especially when I needed help cutting down the word count of my abstract and in designing my poster. I greatly appreciated that with every project in the lab and Science Poster Day, Carole allowed me to start it on my own before she provided guidance. This allowed me to learn a lot about how challenging some elements of research are. I also found the lab meetings this quarter incredibly beneficial for stirring my interests in areas outside just the study I was running. I able to gain great exposure to many different topics in metacognition and memory. I definitely enjoyed Carole's presentation on academic career finding after grad school because it allowed me to consider my own path to grad school and to be aware of the considerations I would have to make before, during, and after graduate school. I also cannot thank Carole enough for recommending me to Prof. Oppenheimer to help instruct Psych 10, which was an invaluable experience. Lastly, I am so excited to help Courtney with the error generation research she will be conducting next year and this paper gave me a chance to engage with the concepts I will hopefully have the chance to help her research.

IMPORTANT NOTICE: The following is an undergraduate research paper submitted as part of the requirements for one quarter of participation in a research apprenticeship in the UCLA Human Perception Laboratory (HPL) in the Department of Psychology at UCLA. This document has not undergone a scientific review and may or may not relate to or correctly represent projects currently or previously undertaken in the HPL. Some features of adaptive and perceptual learning techniques mentioned or described herein are protected by US Patent #7052277 and patents pending. All rights reserved. For information, contact info@insightlearningtech.com.

Active vs. Passive Learning

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 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.

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, the testing effect, to the more specific domain of education using realistic educational materials. 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 in comparison 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).

Evans and Gibbons (2007) further examined the effect of active versus passive learning. 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 learns 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.

References

- Evans, C., & Gibbons, N. J. (2007). The Interactivity Effect in Multimedia Learning. *Computers & Education, 49*, 1147-1160.
- Prince, M. (2004). Does Active Learning Work? A Review of the Research. *Journal of Engineering Education, July*, 223-231.
- Roediger, H. L., & Karpicke, J. D. (2006). Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological Science, 17*(3), 249-255.

IMPORTANT NOTICE: The following is an undergraduate research paper submitted as part of the requirements for one quarter of participation in a research apprenticeship in the UCLA Human Perception Laboratory (HPL) in the Department of Psychology at UCLA. This document has not undergone a scientific review and may or may not relate to or correctly represent projects currently or previously undertaken in the HPL. Some features of adaptive and perceptual learning techniques mentioned or described herein are protected by US Patent #7052277 and patents pending. All rights reserved. For information, contact info@insightlearningtech.com.

Learning And Perception of Music

Erik Shiboski

As an RA in the Human perception lab at UCLA, I've spent the past several months investigating perceptual learning in music. Perceptual learning is defined as the improvement in our ability to pick up information with experience and practice (Gibson, 1969). Recent research suggests that perceptual learning can be systematically produced in vision via perceptual and adaptive learning module (PALM) technology (e.g., Mettler & Kellman, 2014). The goal of our research is to implement PALM technology to explore whether the principles that enhance visual perceptual learning can be extended to abstract auditory pattern recognition of classical music composer styles. As part of this research, I did a literature review of previous research in music perception and learning. This paper presents a series of studies that involve perceptual learning in music and other complex auditory stimuli. I will discuss the implications of their findings, and how our proposed research may expand the results and implications.

In her Tritone Paradox study, Deutsch (1991) discovered individual differences in pitch discrimination that were associated with the language spoken by the listener. The Tritone Paradox is a phenomenon produced when two tones that are related by half an octave (also called a tritone) are presented in succession. Each tone is made up of a series of related harmonics whose amplitudes are shaped by a spectral envelope (a shape that limits the amplitude of various Pitch height refers to the frequency and pitch class refers to the tuning with regards to the equal tempered scale used in Western music (12 available tones in Western music, each with a

different classification). In this study, listeners were asked to judge whether tone pairs formed an ascending or descending pattern. The researchers predicted that culturally acquired representation of pitch classes, and experience speaking a certain language, would affect the perception of these pitch patterns. Two groups of subjects were selected for the experiment and put through a series of tone pair trials. One group was comprised of people born and raised in California while the other group contained people born and raised in southern England. Participants were presented with tone pairs that were matched in amplitude and in randomized order so as not to repeat pitch classes across trials. They were then asked to determine whether the pitch pair was ascending or descending, and responses were recorded. Two sessions were administered and data was averaged over both.

The results from both sessions demonstrated a clear relationship between the cultural background of the subjects and their judgements of tone pairs. Representation of pitch classes seems to influence perception of musical patterns and speech. The pitch pairs that Californians tended to judge as ascending, were generally judged as descending by Brits. Both groups showed different response patterns across different pitch pairs. The results offer strong support for the theory that individuals acquire mental representations of a pitch class circle (pitches presented around a circle, ordered chromatically) during development with a particular orientation with respect to height of the pitches. Peak pitch classes in this mental representation (those that are at the peak of the mental pitch-class circle) are derived from exposure to speech sounds, and class representation, and peak pitch-classes are a possible example of auditory perceptual learning facilitated by repeated exposure to speech sounds in development. This could imply that repeated exposure to certain pitch patterns or speech patterns could train for pitch-class discrimination or other sorts of pattern recognition in auditory stimuli. Although the results are encouraging, it

remains to be seen if perceptual learning of more complex musical patterns could be trained.

In their paper “Conceptual And Motor Learning In Music Performance”, Palmer and Meyer (2000) investigate the differences in motor skills and learning among expert and novice music performers. Previous research has shown that among experts, necessary motor movements are independent of mental plans for action (hence the term motor-independence). The researchers used a transfer-of-learning task to demonstrate these motor skill differences. These tasks involve transferring movement control learned in one situation to a novel task. They predicted that in this case, expert music performers would demonstrate motor-independence when playing novel pieces. For the first experiment, researchers recruited 16 pianists with a mean of 18 years of piano playing experience (expert pianists). They were given a simple melody to learn and then several transfer trials that varied in terms of motor information (different hand/finger positions) or pitch information (different pitch sequence), or both. Results were recorded from a computer monitored electric keyboard that recorded key-press onsets and offsets and compare it to the notated musical sequences. The same procedure (with slight alterations in the learning phase) was completed with novice pianists (children with an average of 4 years of piano experience) discovered that experts have a dissociation between the movements required to produce a musical sequence and the mental representations for the performance. This “dissociation” refers to an observable disconnection between the mental plans and representation of a musical sequence and the actual movements used to perform it. This implies that participants do not have to map movements onto the mental representation of the sequence. The movements are executed in an autonomous way. Experts perform the test variations of the learned melody with the same keypress onsets and durations as the learned melody, regardless of whether motor information or pitch information was varied. Even in experiment 2, children with slightly higher skill levels

showed this dissociation across transfer tasks. The results indicate a form of perceptual learning in music performance: motor independence, and increased ability to transfer learning across novel melodies is the result of improvement in the ability to differentiate motor and conceptual aspects of the music. This perceptual learning is likely the result of years of practice. It remains to be seen how this transfer of learning could be observed in other scenarios. How does learning transfer across musical genres? And how might one learn to distinguish elements of style within musical excerpts?

In a study exploring implicit learning of tonality, Tillmann and colleagues (2000) investigated the implicit knowledge acquired by exposure to different combinations of tones in the environment. Much in the same way humans learn language and acquire a mental representation of its structure, the “grammar” and structural elements relating pitch, chords, and key areas are implicitly acquired and represented. In order to investigate the implicit knowledge of this musical grammar, the researchers designed a hierarchical self organizing model (HSOM) various tonal inputs. The structure of this model was very similar in structure to those used for modeling language perception. The networks in the model are initially unorganized but by repeated exposure to musical stimulus, relationships are self-organized topographically across the layers of the model (layers representing different levels of ‘grammar’ within the western music system). After training the model, each unit becomes specialized to a particular input pattern, and units are topographically organized such that similar input patterns activate nearby units. Using this system, the researchers trained several models using simple and complex harmonic material (simulating passive input of tonal information). They then compared the learned models to human participants in a variety of experiments on perception of tonality (using data from previous experiments with human subjects). Simulations were completed with the

learned models in chord priming, key modulation, chord relations, and other components of musical perception and knowledge.

Given the complexity of relationships within various aspects of music (key centers, scales, chords, and pitches), one could expect that perceiving Western musical structures requires substantial explicit knowledge of the rules and systems. But the experimental data observed from the simulations completed with these HSOMs, along with data from human participants, shows that explicit learning is unnecessary to develop knowledge of musical structures. The data suggests that passive exposure to music can create a complex understanding of musical ‘grammar’. For this sort of understanding to develop, there must be some level of perceptual learning resulting from repeated exposure to musical stimulus. The patterns that relate chords, keys, and pitches within Western music become easier to extract. In this way, an average listener scale. If these complex elements of music can be learned implicitly, then what other elements of music might we acquire knowledge for implicitly? Could listeners be trained to acquire knowledge of structural elements of music by a combination of passive and active (asking questions) exposure to music? Our research will begin to explore whether perceptual learning of music can be systematically produced with training using learning technology.

In our proposed study, we will be implementing perceptual and adaptive learning module (PALM) technology that trains for the recognition of 4 classical composers based on their composition style. We selected composers from the Baroque and Romantic periods, and used only piano music selections. The series of software-based learning trials will involve comparing musical excerpts (audio clips), matching music to a composer’s name (via multiple answer options presented onscreen), and receiving feedback for responses. Subjects will progress through the PALM based on accuracy, and the module will dynamically calculate the

participant's level of mastery throughout. To complete the module, a participant must reach a specified level of mastery for each composer. The study will also test, before and after the series of PALM learning sessions, subjects' ability to recognize novel clips from learned composers. Tests will involve matching a music clip to the composer's name, with a total of 7 choices for each trial: the four trained composers' names, "Other Baroque", "Other Romantic", and "Other Period". Novel clips and composers will be introduced during the tests to control for memorization of the excerpts and to see if participants can distinguish what was trained (composers, periods) from what wasn't.

We hypothesize that the PALM will improve the subjects' ability to extract relevant features and patterns to classify the style of the selected composers. The previous research discussed above indicates several ways in which knowledge of tonality and musical structure can be learned through perceptual learning. We hope to extend the results of this research by exploring perceptual learning of composers' styles, a highly abstract element of composed music, and by creating a module that trains participants in this sort of learning. Our research could demonstrate that perceptual learning technology can train for abstract patterns in auditory stimuli, and ultimately offer important implications for perceptual learning in education. The training of verbal language is a real-world example of dynamic pattern recognition. Our research could offer applications for the training and learning of verbal language, and other dynamic stimuli (visual or auditory). By extending the implementation of perceptual learning to more abstract elements of dynamic stimuli, we hope to broaden the educational implications of research in perceptual learning.

References

- Deutsch, D. (1991). The tritone paradox: An influence of language on music perception. *Music Perception*, 335-347.
- Palmer, C., & Meyer, R. K. (2000). Conceptual and motor learning in music performance. *Psychological Science*, 11(1), 63-68.
- Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: a self-organizing approach. *Psychological review*, 107(4), 885.

Analogical Reasoning with Rational Numbers

Eugene Goh

Abstract

Previous studies have shown that mathematical and real-world relations guided people's ability when solving arithmetic word problems and when constructing algebraic expressions. We propose that rational numbers can serve as relational models. A fraction's structure (a/b), also known as its bipartite structure, encourages estimation or counting of the size of two sets. In particular, fractions are a more accurate counting strategy when the sets comprise of discrete elements. Discrete elements refer to elements that are not continuous. On the other hand, decimal notations encourages a one-dimensional estimate of a ratio, which typically leads to a procedure less accurate than counting, which is the procedure that fractions encourage. We constructed analogical reasoning problems that required subjects to map rational numbers (fractions or decimals) onto displays, which either depicted part-to-whole or part-to-part ratios between two quantities. Furthermore, we even varied the nature of the displayed quantities, which could be discrete, continuous, continuous but parsed into discrete components. When reasoning about discrete elements, fractions were more accurate than decimals. However, when quantities were continuous, accuracy was lower for both number types, with decimals being more advantageous. These findings support the general conception of mathematical thinking as a form of modeling, where mathematical formats mediate between the structure of the real world and quantitative concepts.

Analogical Reasoning with Rational Numbers

Our relational models based on our analysis of fractions leads us to believe that, fractions allow more accurate reasoning about bipartite relational structures, even though decimals are more effective in conveying one-dimensional magnitudes than fractions are. We created pictorial displays of a set comprised of two subsets, paired with either a fraction or decimal value representing a certain ratio relation within the display in order to evaluate whether fractions are superior to decimals as relational. The ratio relationships were a *part-to-whole* or a *part-to-part* ratio. Numerical values were paired with displays that showed two subsets that were either broken down into countable units or shown together as one continuous mass.

When discrete units were provided, counting is a likely strategy, which should generate accurate measures of subsets that align directly with the numerator and denominator of a fraction. However, if counting of subsets/sets were used for decimals, then additional processing are required to translate these results into decimal form. Alternatively, decimal magnitudes may be estimated directly (Jacob et al., 2012). However, estimation is likely to be less accurate than counting, resulting in more errors when decimals, rather than fractions, are paired with displays of countable items. For continuous displays, the bipartite format of fractions may still encourage counting (e.g., by mentally slicing the display into units). However, accuracy is likely to be sacrificed, reducing or eliminating the advantage of fractions over decimals.

We predicted that for displays of countable entities, fractions would yield greater accuracy than decimals in both relation identification and analogical reasoning.

Method

Participants

We had 52 undergraduate students from the University of California, Los Angeles (mean age = 21; 30 females). Participants were randomly assigned in equal numbers to the two between-subjects conditions. We provided participants with course credit for participating in the study.

Design

The study was designed to be a 2 (number type: decimals vs. fractions) X 2 (relation type: part-to-part vs. part-to-whole ratios) X 3 (display type: continuous, discretized, discrete) study. The relation and display types were within-subjects factors, whereas the number type was a between-subjects factor.

Materials

Discrete items included circles, squares, stars, crosses, trapezoids, and cloud-like shapes. Continuous items included displays of rectangles that differed in width, height, and orientation (both vertically and horizontally). Discretized items were identical to the continuous items, except for the fact that in the discretized case, the rectangles were divided by dark-lines into equal-sized objects. Figure 1 below shows all three of the display types. The red and green colored items depict the stimuli used in actual test trials, whereas yellow and brown colored items were used in practice trials. We varied the color such that it represented both the larger subset and smaller subset. Half of the trials had the red subset as the numerator, and the other half had the green subset as the numerator. We used Superlab 4.5 on Macintosh computers to display these stimuli. Participants' response times and accuracy were recorded.

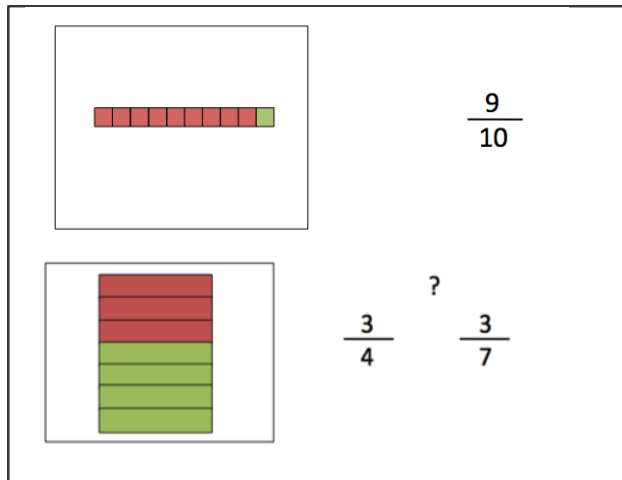


Figure 1. Examples of analogy problems.

We paired discrete, discretized, and continuous displays with either a fraction or decimal that represented a part-to-part ratio or part-to-whole ratio. Participants needed to identify which of the two numbers in the target correctly matched the display with the same relationship specified in the source. The task required making a choice of the correct number to complete the target analog. Number type was always the identical across the source and target. Solving an analogy problem required a participant to first identify the ratio relation in the display characterized by the number given. Once this higher-order relation was extracted from the display, participants had to identify the same relation type in another target display.

Procedure

Participants read and signed a consent form upon arriving at the experiment room. Participants were then quickly seated in a computer room after the experimenter was done setting up. Participants were told to try to go as quickly as possible without sacrificing accuracy. After reading instructions and completing 12 practice trials with feedback, participants proceeded to complete 72 test trials. Participants first saw the source problem and were asked to identify the relationship between display and value. Participants hit the space bar when they identified the

relationship. After the space bar was pressed, the target was shown on the screen below the source, so that both the source and target were on the screen at the same time. Participants had to choose which of two numbers shared the same relationship as the display type shown by the source. The z and m keys were labeled with “L” and “R”, so that participants could choose which display they wanted. The z key corresponded to the left display, whereas the m key corresponded to the right display.

Results

A 3(display: continuous, discretized, discrete) X 2 (relation type: PPR vs. PWR) X 2 (number type: fraction vs. decimal) mixed factors ANOVA was used to assess differences in response time (RT) and accuracy. Only accuracy and mean RT on correct trials were computed for each participant in each condition. There was no significant main effect of relation types on either measure (accuracy: $F(1, 50) = 2.58, p = 0.11$; RT: $F(1,50) = 0.73, p = .40$).

Looking at Figure 2, there appears to be an interaction between display and number type. The analysis revealed a significant interaction between display type and number type $F(2, 49) = 20.59, MSE = 1.8, p < .001$. Planned comparisons indicated that accuracy was higher for fractions than decimals in discrete condition ($F(1, 50) = 28.96, MSE = 7.3, p < .001$ and discretized condition $F(1, 50) = 10.06, MSE = 8.3, p = .003$. However, accuracy did not differ across the two number types for the continuous condition $F(1, 50) = 0.86, MSE = 7.7, p = .36$.

Looking at Figure 3 In analyzing RTs, we found a significant interaction between number type and display type $F(2, 49) = 16.19, MSE = 2721, p < .001$. Planned comparisons revealed that RTs were slower with decimals than with fractions for the discrete condition $t(50) = 2.70, p = .01$. However, RTs had a strong trend for the discretized condition $t(50) = 1.87, p = .07$.

Finally, for the continuous conditions, RTs for fraction and decimals did not reveal a significant difference $t(50) = 1.45, p = .15$.

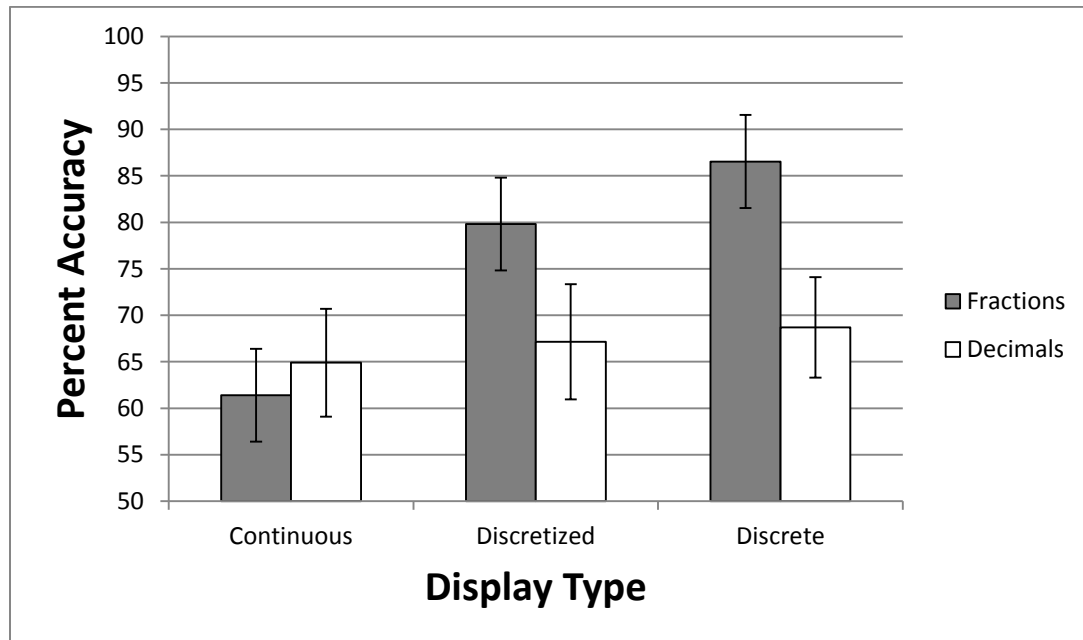


Figure 2. Mean Accuracy of analogical inferences using fractions and decimals across different types of displays.

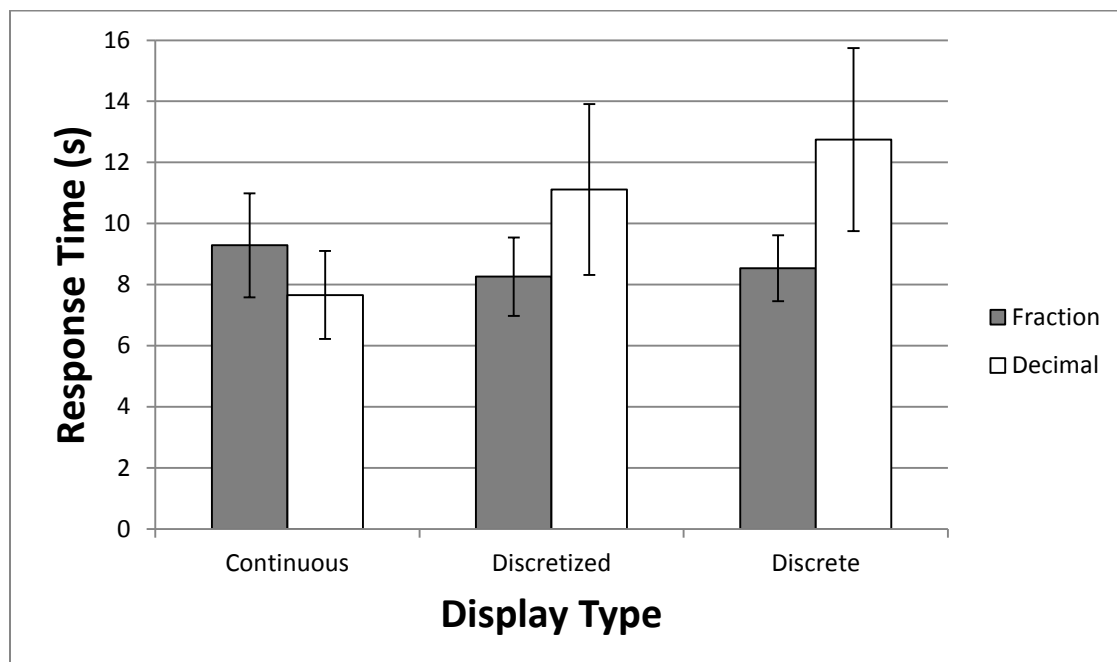


Figure 3. Mean response time for analogical inference using fractions and decimals across different types of displays.

Discussion

The findings of this study showed that fractions are much better suited for discretized and discrete displays, both in terms of accuracy and RT. Fractions were shown to be advantageous for identifying ratio relationships when these ratios were represented by both fractions and decimals. DeWolf et al (2013) assert that fractions are better for representing the relations between two distinct sets because of their bipartite structure. In this experiment, the visual displays might have been a limiting factor as these results were shown only with these displays. Ratios were evaluated more accurately when the number was represented as a fraction rather than a decimal when displays conveyed countable entities, which consist of either sets of discrete objects or discretized objects. However, continuous quantities showed no such advantage for fraction or decimals. Accuracy was similar in evaluating ratio relations, with decimals being slightly more accurate, but not significant enough to be considered advantageous. RT was also slightly faster for decimals than fractions for the continuous displays, but there was no significant advantage there as well.

These findings imply that people are more quickly to identify a number that correctly maps to a ratio relation when the internal structure of the number can be mapped to a fraction. The fraction advantage extends beyond the identification of ratio relationships, as fractions facilitate mapping of higher-order relations when the specific quantities differ. A possible explanation as to why people performed worse on continuous displays might be that adults exhibit misconceptions of the complex conceptual structure of fractions (Siegler et al., 2011; 2013; Ni & Zhou, 2005; Stigler et al., 2010). Another possible explanation is that students often

find it difficult to understand how whole numbers that are within the fraction contribute to its overall magnitude (Ni & Zhou, 2005; Vamvakoussi & Vosniadou, 2010). This fraction advantage reflects the fact that fractions align well with discrete quantities. On the other hand, decimals performance was relatively equal across conditions, and less accurate for all quantity types than fractions, which suggests that decimals are evaluated using estimation rather than counting.

References

- DeWolf, M., Grounds, M. A., Bassok, M., & Holyoak, K. J. (2013). Magnitude comparison with different types of rational numbers. *Journal of Experimental Psychology: Human Perception and Performance*. Advance online publication. doi: 10.1037/a0032916
- Ni, Y., & Zhou, Y. D. (2005). Teaching and learning fraction and rational numbers: The origins and implications of the whole number bias. *Educational Psychologist*, 40, 27–52.
- Siegler, R. S., Thompson, C. A., & Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive Psychology*, 62, 273-296.
- Siegler, R. S., Fazio L.K., Bailey, D.H., & Zhou, X. (2013). Fractions: The new frontier for theories of numerical development. *Trends in Cognitive Sciences*, 17, 13-19.
- Stigler, J. W., Givvin, K. B., & Thompson, B. (2010). What community college developmental mathematics students understand about mathematics. *The MathAMATYC Educator*, 10, 4-16.
- Vamvakoussi, X., & Vosniadou, S. (2010). Understanding the structure of the set of rational numbers: A conceptual change approach. *Learning and Instruction*, 14, 453- 467.

IMPORTANT NOTICE: The following is an undergraduate research paper submitted as part of the requirements for one quarter of participation in a research apprenticeship in the UCLA Human Perception Laboratory (HPL) in the Department of Psychology at UCLA. This document has not undergone a scientific review and may or may not relate to or correctly represent projects currently or previously undertaken in the HPL. Some features of adaptive and perceptual learning techniques mentioned or described herein are protected by US Patent #7052277 and patents pending. All rights reserved. For information, contact info@insightlearningtech.com.

Effects of Active and Passive Learning on Cognitive Load

Giancarlo Sanguinetti

Learning can be defined as the modification of behavior as the result of the acquisition of new knowledge, whether it be through experience or discovery. Learning can be applied to nearly any conceivable domain of knowledge, and can even occur without any conscious awareness of it (Kuldass, Ismail, Hashim, & Bakar, 2013). The majority of learning takes place in a controlled, educational setting, but even then a variety of methods can be used to relay the same information. The aim of this paper is to understand the strengths and weaknesses of two types of learning used in modern curricula: active learning and passive learning. Active and passive learning differ in the amount of interactivity with which the learner engages the material. To frame these observations and extrapolations, I will discuss research on this topic by focusing on a necessary aspect of learning and human cognition in general, called cognitive load. Cognitive load is a psychological concept that describes the allocation of mental resources, such as attention or memory capacity, usually in terms of problem solving. The final result of this paper is to provide a wealth of evidence that demonstrates that both active and passive learning have their respective advantages, and with enough awareness of the factors that affect cognitive load the best method can be selected to maximize learning in a variety of educational curricula.

The majority of passive learning is conducted by having the learner read the material and study it, using common methods such as rehearsal. In domains such as mathematics and science, passive learning would involve the use of “worked examples”, which are problems that are

presented along with their solutions, and the necessary steps in between. Sweller (2006) performed an analysis of several studies that employ the use of worked examples, establishing a number of principles that expound on the several approaches used by the human mind, a natural information processing system. The five principles are: information store, borrowing, randomness as genesis, narrow limits of change, and the environment organizing and linking principle. The information store principle refers to the fact that the majority of a human's cognition is dominated by the content knowledge accumulated through a lifetime of learning. The borrowing principle builds on the information store principle by explaining that the majority of learned knowledge is a perpetual construction between knowledge gained through self-experience and knowledge that is "borrowed" from others via learning; it relates human cognition as a constant process of assimilating borrowed knowledge. The principle of randomness as genesis explains that for the gaps of information knowledge, new information must be created. This principle is most directly related to problem solving, as the information that is "created" is more aptly described as a hypothesis generated by the learner, synthesized from the previous two principles, which will be tested by trial-and-error. The narrow limits of change principle describes how information in long term memory is only changed in small increments from the previous two principles because of the limited capacity of working memory. Lastly, the environment organizing and linking principle states that working memory serves as a sort of bridge between new knowledge and long term knowledge, where the new knowledge is "organized and linked" to related content knowledge. Sweller goes on to use these principles as possible explanations for why worked examples are less taxing in terms of cognitive load compared to examples that require problem solving. For instance, because a learner's cognitive load is lower for studying passively, they have more resources to dedicate committing that

information into long term memory. Active learners have an increased cognitive load because their attention must be divided between generating answers, processing feedback, and committing the same amount of information to memory, or more. Arguably, because the active learners get feedback, they should improve their problem solving skills overtime, leading to a reduced cognitive load over time. At the testing phase, active learners should perform just as well, if not better than passive learners simply because their problem solving ability is more refined, such that hypothesis generating becomes a trained skill that stays with the learner longer than simply committing to memory a number of examples that might decay in long term memory (Catrambone & Yuasa, 2006).

As explained by Gerjets, Scheiter, & Catrambone (2004), cognitive load can be manipulated in a variety of manners, seeing as the differences between the types of cognitive load, intrinsic, germane, and extraneous, are based on different aspects of the presented material. For clarification, intrinsic cognitive load refers to the inherent difficulty of understanding the material, such as algebra having less of an intrinsic cognitive load than multivariable calculus, the latter being more complex than the former. Extraneous cognitive load refers to the factors that affect how the material is presented, such as trying to teach someone what a square is; while it can be described in terms of geometrical relationships, it is less effort to simply draw a square. Germane cognitive load refers to the cognition processes used by the learner to engage the material; in essence, it describes how much effort a learner expends studying the material, such as relating it to information already known. Germane cognitive load contributes to learning the most. In conjunction with the principles of information processing (Sweller, 2006), suggestions for reducing intrinsic cognitive load in active learning can be providing an apt model of translation on which the learner already has strong content knowledge,

thereby using the borrowing principle instead of the randomness as genesis principle, and effectively strengthening working memory by utilizing mental resources already in place, such as learning a new programming language made easier when described in terms of a programming language already well known (AlSagheer, 2011). Contrarily, both extraneous and intrinsic cognitive load can be reduced by passive learning working with worked examples, as demonstrated with much of the aforementioned research, seeing as it would only employ the borrowing principle and the principle of environment organising and linking principle, both of which are significantly less taxing on working memory.

An additional example of previous research that explores the relationship between problem-solving with examples that were either engaged passively or actively and cognitive load can be seen in the work of Paas and Van Merriënboer (1994), who examined the two methods of learning in the domain of geometry. The applicability of this research is immediately relevant as geometry is considered a standard of education in any modern curricula. The experimental conditions of this study separated participants by the type of practice, either working problems conventionally (active) or studying worked examples (passive), and the variability of the practice, or how similar the questions were to each other in terms of difficulty. In addition to performance on both a training phase and a test phase, measures included a survey of self-estimated of cognitive load and physiological measures, such as heart rate recorded by an ECG, by participants during the latter phase of the study. Results demonstrated that the passive learning condition had better test performance and rated their cognitive load significant lower than the active learning condition, which was corroborated by the physiological measures. In addition, participants whose questions were high in variability showed better scores than their low variability counterparts. This can be explained by previous research that reveals that

providing more varied examples to participants creates a “generalizability effect”; by providing a wide range of examples that fall within the boundaries of a category selection, the learner generates an array of exemplars to draw from at testing (Wulf & Schmidt, 1997).

Sweller’s principles can be used to more closely analyze in the differences between active and passive learned using research conducted by Markant and Gureckis (2014). In this study, the idea of hypothesis testing was examined through either active learning or passive learning, denoted as either selection or reception, respectively. The researchers reviewed an extensive amount of literature on the topic, and based their hypothesis on the fact that the differences in the complementary methods of learning, active learning and passive learning should result in identical patterns of results at testing, but each method would have distinct advantages in certain areas. They assumed that “learning via selection would be more efficient because it generates more informative data” that “may lead to a training experience that is uniquely optimized to refine their existing knowledge”. Conversely, they acknowledged previous data showing that when the rules for learning were considered more disjunctive than simple, passive learning lead to stronger results (Schwartz, 1966). The experiment consisted of a category learning task separated by either a rule-based learning method or an information integration method, the difference between the two being how many dimensions to which the category criterion was judged. For example, if the category consisted of matching a target color to one selected from a palette, then the rule based category would use one dimension, such as brightness, whereas the information integration category would use multiple dimensions, such as hue, saturation, and brightness. Within these conditions the participants were either designated to selective or receptive learning, and finished with a post-test of category judgments. Their results show distinctly that selection learners would outperform receptive learners, most significantly on

the rule-based condition. Reasons for this might be the same as discussed above, as having to increase germane cognitive load through problem solving facilitates a deeper understanding of the material (Gerjets, Scheiter, & Catrambone, 2004). Differences in the conditions could be a result of the increased intrinsic load of the information integration, as needing multiple dimensions to categorize stimuli result in complex processing of both the principles of borrowed knowledge and randomness as genesis.

The culmination of the research presented in this paper helps to frame the following arguments. Anytime educational material is designed, it is necessary to consider the level of mastery of the target demographic before instruction and the goals of the instruction. If the material is inordinately complex to a novice, then a passive curriculum is recommended, especially in topics that require a strong proficiency in the perceptual aspects of the problem, as passive learning provides a wealth of exemplars that can teach the learner the range of boundary conditions. Active learning challenges the learner to generate a number of possible answers and select one that most closely matches the exemplars they know to be true (Catrambone & Yuasa, 2006; Markant & Gureckis, 2014). If long term knowledge mastery is the goal, then for the majority of knowledge domains, active learning should be heavily considered as despite the increased demand in cognitive load, the manipulation of the material fosters a deeper understanding of the material. While both active and passive learning should generate similar results after an immediate post-test (Swinnen, 1990), a delayed post-test would be indicative of the superiority of active learning, possibly because the act of constantly generating hypotheses lingers even if specific exemplars are forgotten.

More evidence of the factors that the design of the studying materials are affected in a variety of ways can be observed in the studies whose results demonstrate that participants who

engaged in material that definitely pushed the boundaries of categorization learned more efficiently than those who did not (Markant & Gureckis, 2013). In addition, participants who had to use little of their working memory resources because of the ease of instruction did just as well, if not better than participants who engaged in deeper understanding of materials even if they indicated a high cognitive load (Paas & Van Merriënboer, 1994). If the studying material is complex enough to warrant both “rule-based and knowledge-based exemplar generation” then passive learning conditions help “direct attention to goal-relevant aspects of the task...preventing capacity-demanding actions”.

Using the knowledge generated by the discussed research, we can apply it to more specific domains, and see if active and passive learning have definitely better distinctions in specific domains of knowledge, such as active learning being better for the physical sciences while passive learning being more suited for perceptual heavy domains, such as art history. In addition, we can apply this research to different areas of study that would be immediately relevant to modern education settings, such as multimedia learning or perceptual learning. Future directions for educational psychology can incorporate the methods discussed here to provide a reliably standardized model of education, one that can be adapted to modern needs, and provide populations with the skills and knowledge necessary for a well-rounded education, such that problem solving for more abstract tasks can be better answered from knowledge generated from a number of modalities and domains.

References

- AlSagheer, A. (2011). The Transfer of Learning: Exploration of Benefits and Perspectives in the State of Kuwait. *International Journal of Managment and Information Systems*, 15(2), 39-48.
- Catrambone, R., & Yuasa, M. (2006). Acquisition of procedures: The effects of example elaborations and active learning exercises. *Learning and Instruction*, 16(2), 139-153.
- Gerjets, P., Scheiter, K., & Catrambone, R. (2004). Designing Instructional Examples to Reduce Intrinsic Cognitive Load: Molar versus Modular Presentation of Solution Procedures. *Instructional Science*, 33-58.
- Kuldas, S., Ismail, H., Hashim, S., & Bakar, Z. (2013). Unconscious learning processes: mental integration of verbal and pictorial instructional materials. *Springer Plus*, 92-105.
- Markant, D., & Gureckis, T. (2014). Is It Better to Select or to Receive? Learning via Active and Passive Hypothesis Testing. *Journal of Experimental Psychology: General*, 143(1), 94-122.
- Paas, F., & Van Merriënboer, J. (1994). Variability of Worked Examples and Transfer of Geometrical Problem-Solving Skills: A Cognitive-Load Approach. *Journal of Educational Psychology*, 86(1), 122-133.
- Schwartz, S. H. (1966). Trial-by-analysis of processes in simple and disjunctive concept-attainment task. *Journal of Experimental Psychology*, 72, 456-465.
- Sweller, J. (2006). The worked example effect and human cognition. *Learning and Instruction*, 16, 165-169.

- Swinnen, S. (1990). Interpolated activities during the knowledge-of-results delay and post-knowledge-of-results interval: Effects on performance and learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 692-705.
- Wulf, G., & Schmidt, R. (1997). Variability of Practice and Implicit Motor Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(4), 987-1006.

IMPORTANT NOTICE: The following is an undergraduate research paper submitted as part of the requirements for one quarter of participation in a research apprenticeship in the UCLA Human Perception Laboratory (HPL) in the Department of Psychology at UCLA. This document has not undergone a scientific review and may or may not relate to or correctly represent projects currently or previously undertaken in the HPL. Some features of adaptive and perceptual learning techniques mentioned or described herein are protected by US Patent #7052277 and patents pending. All rights reserved. For information, contact info@insightlearningtech.com.

Adaptive and Fixed Schedules in Complex Fact Learning

Jason Ku

In the domain of memory and learning, numerous studies have shown that spaced retrieval practice leads to better long-term retention than massed practice. Studies have shown that the amount of spacing between repeated tests of a certain item also makes a significant difference in learning. Some spacing sequences include equally spaced retrieval sequences, fixed expanding retrieval sequences (T.K. Landauer & R.A. Bjork, 1978), random sequences and adaptive sequences (Atkinson, 1974; Mettler & Kellman, 2010). Our current study compares the effectiveness of fixed expanding sequences with that of adaptive sequences.

A fixed expanding sequence is one where the initial spacing between repeated tests is very small and gradually increases with subsequent tests (T.K. Landauer & R.A. Bjork, 1978). The model was based on the idea that increasing the difficulty of testing, by manipulating the spacing between tests, increases long-term retention. Adaptive sequences, originally developed by Atkinson (1974) used an algorithm based on response accuracies and overall performance history to determine the spacing. Mettler, Massey, and Kellman (2011) improved upon Atkinson's model by factoring in response time as well as accuracy. Their adaptive algorithm, called Adaptive Response Time Based Sequencing—ARTS, attempts to optimize spacing between presentations by increasing or decreasing the delay based on how quickly and accurately an answer is made.

One particularly important concept when comparing adaptive and fixed sequencing is the idea of item independence. When items are independent of each other, there are no relationships between the items and learning one item does not help nor interfere with the learning of another item. Simple memory experiments, such as one where participants memorize the capitals of different states, exhibit item independence. For example, learning that the capital of California is Sacramento does not help one learn what the capital of Texas is. Independent items are also not easily confused during retrieval since there are no relationships between the items.

Fixed schedules of spacing assume the independence of items. In fixed schedules, all items have approximately the same spacing between trials, which assumes that all items are at about the same level of difficulty (T.K. Landauer & R.A. Bjork, 1978). This may become problematic when there are some relations between items that make certain items more or less difficult. In the states and capitals example, there are a few cases where this can be observed. Suppose a participant first learned the capital of North Carolina, then tried to learn the capital of South Carolina. This may cause the participant to confuse their capitals since there are similarities in their names. In this case, the items are said to be dependent. When there are relations between items, learning of one item may make it more difficult to learn another item. Varying levels of difficulty require different amounts of spacing to account for the item differences. A fixed schedule does not account for item variations.

Adaptive schedules however, are designed to be sensitive to differences in learners as well as variations between items. Since adaptive sequences are generated dynamically based on a learner's reaction time and accuracy, it does have a built-in mechanism for detecting individual item differences. This may be useful when there are many complex relationships between items being learned.

In this current experiment, participants completed a chemistry nomenclature module in which they learned the names and formulas of various chemical ions. The module employed either a fixed continuously expanding sequence (T.K. Landauer & R.A. Bjork, 1978) or an adaptive sequence (Atkinson, 1974; Mettler & Kellman, 2010). The purpose of this research is to determine which type of spacing results in greater gains in learning.

In the domain of chemistry, chemical names are highly related to one another which make them easy to confuse. Suppose a participant learned the following formulas: NO_3^- = Nitrate, NO_2^- = Nitrite, N^{3-} = Nitride, ClO_3^- = Chlorate, ClO_2^- = Chlorite Ion, Cl^- = Chloride. The participant may try to distinguish some patterns that help them memorize the formulas. For example, they may conclude that the –ate suffix corresponds to ions with three oxygen atoms, that the –ite suffix corresponds to ions with two oxygen atoms and the –ide suffix denotes ions with no oxygen atoms. While this general rule may help with the learning of many ions, there are cases where the rule does not apply and using the rule may interfere with learning in these special cases. One such case is sulfur: SO_4^{2-} = Sulfate, SO_3^{2-} = Sulfite, S^{2-} = Sulfide. Applying the rule here would result in confusing the sulfite ion for sulfate since generally ions with three oxygen atoms have the –ate suffix.

When conducting the analysis of the results of this experiment, we were interested in whether or not we could see the relationships between items. We plotted a confusion matrix to see where the most common errors were made. Figure 1 shows the responses of participants on the y-axis and the correct answers on the x-axis. The color of the boxes indicate the frequency of a particular response. The majority of responses fall on the diagonal where the participant gave the correct response. However, there are several areas where participants were making frequent mistakes. Figure 2 displays the items that were most frequently confused. These items seem to be

of the type discussed earlier with many inter-item relationships. Figures 3,4 and 5 display the actual frequency of confusion averaged across participants.

One interesting trend to notice is that the frequency of confusions are not symmetrical across items. Figure 5 shows the frequency of confusion for sulfite and sulfate. Students were much more likely to incorrectly answer sulfate than sulfite. One possible explanation is that when there is a possible confusion, students rely on what is most familiar to them when making their response. Since sulfate is a much more commonly known chemical, it is more easily retrieved. This creates yet another variable to consider when generating spacing sequences. That is, how do the familiarity of items affect the learning of related items?

We were also interested in seeing if these confusions were reflected in measures of learning strength. Specifically, we wanted to see if the reaction times differed between the confusable and non-confusable items (free-response questions only). We compared the average reaction times of the 10 most frequently confused items with the average reaction times of all the other items in the set (about 15 items). Figure 6 shows the average reaction times for correct and incorrect responses of the two types of items. When the students made a correct response, the reaction times were about the same on average. The mean reaction times for incorrect responses also did not differ significantly. Interestingly however, the student's seemed to respond slightly slower for the non-confusable items. We expected the confusable items to take longer since students would require more time to think about the answer. Perhaps when answering easily confusable questions, students were actually more confident and not even aware that they are confusing the items. For the incorrect non-confusable items, students were less confident and took more time to think about the answer. From these results, it is unclear whether confusions are reflected in measures of learning strength.

The primary question we wanted to answer when conducting these analyses was whether or not adaptive schedules account for the difficulties with confusable items. An easily confusable item is generally more difficult to learn than a distinct, non-confusable item. More difficult items require less spacing to provide more frequent exposure to that item. If the delay for a difficult item is stretched out too far, students may forget the item and be more likely to make a mistake. Thus, if the adaptive schedule did account for the increased difficulty of confusable items, we should be able to see a noticeable average delay difference between the two types of items. Figure 7 shows the average delay between presentations of confusable and non-confusable items. Though the difference in delay is not significant, the difference is in the right direction. The confusable items had a slightly shorter delay than the non-confusable items.

Adaptive algorithms do have a mechanism for detecting variations between learners and item differences, but it is unclear whether the current algorithm correctly accounts for the differences. The currently algorithm only takes reaction time into account when the correct answer is made. An incorrect response automatically results in an enforced 1-space delay and the reaction time of the incorrect response is ignored. Perhaps an updated algorithm could use the reaction time of incorrect responses as an additional variable. It is clear from figure 6 that there is a noticeable difference in reaction times between confusable and non-confusable items for incorrect responses. It could be the case that a shorter reaction time for incorrect responses indicates confusion and the algorithm could be modified to account for the confusion.

References

- Atkinson, R. C. (1974). *Adaptive instructional systems: Some attempts to optimize the learning process*. Stanford University, Institute for Mathematical Studies in the Social Sciences.
- Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625-632). London: Academic Press.
- Mettler, E., & Kellman, P. (2010). Adaptive sequencing in perceptual learning. *Journal of Vision*, 10(7), 1098-1098.
- Mettler, E., Massey, C., & Kellman, P. J. (2011). Improving adaptive learning technology through the use of response times. In *Proceedings of the 33 rd Annual Conference of the Cognitive Science Society*. Boston, MA: Cognitive Science Society (pp. 2532-7).

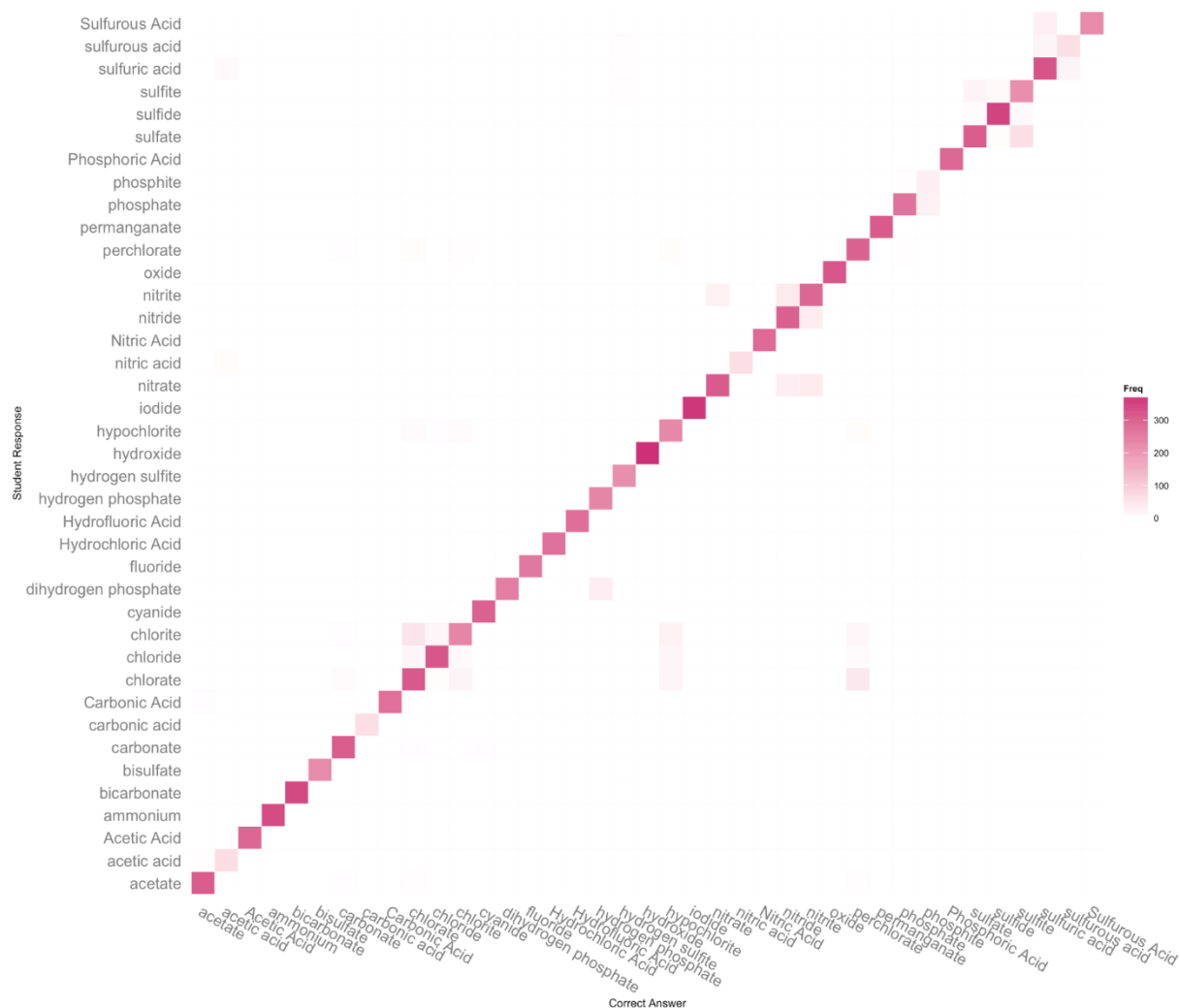


Figure 1. Confusion Matrix of student's responses plotted against the correct answer. Subset of complete confusion matrix with only valid answers displayed.

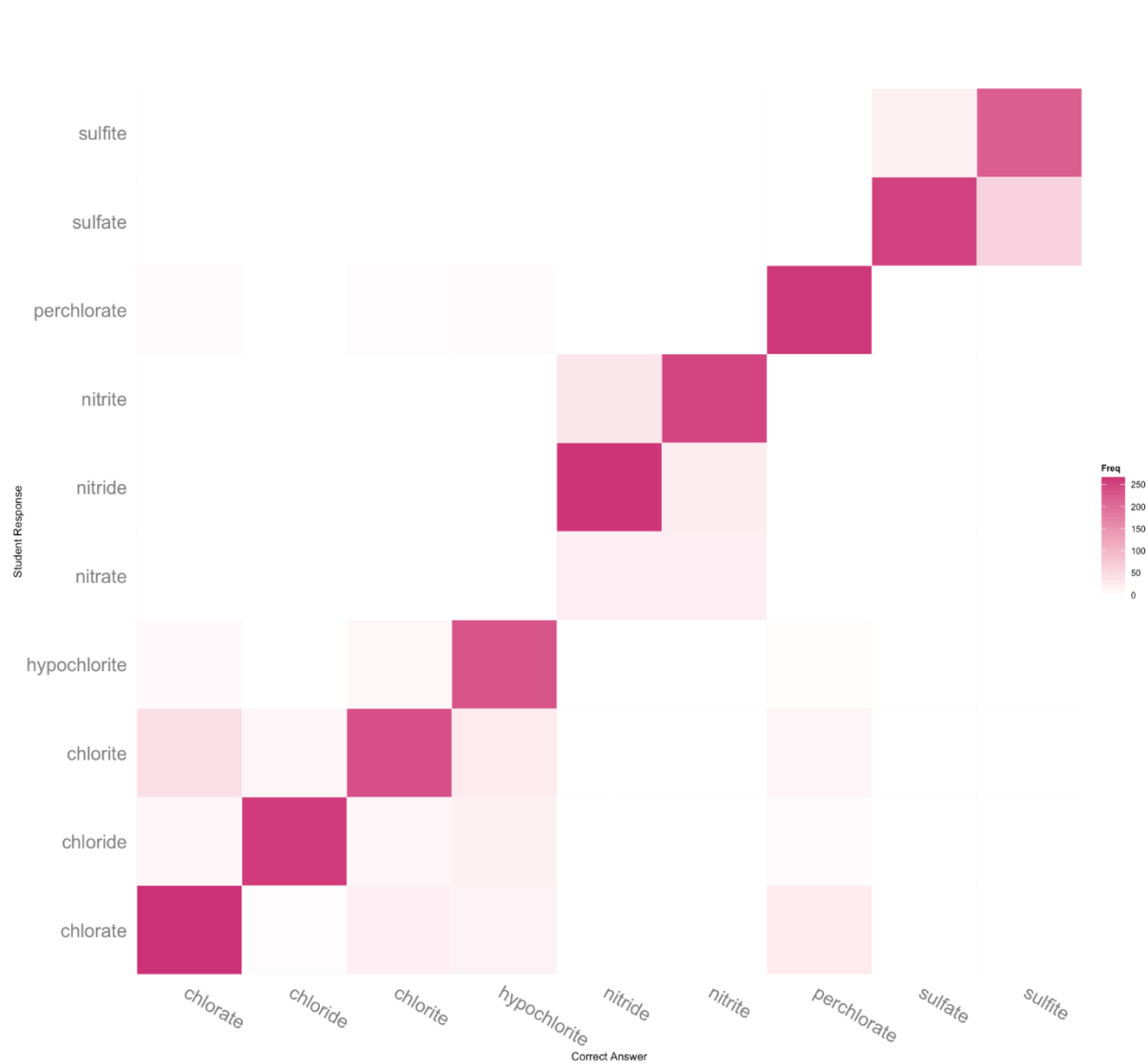


Figure 2. Confusion matrix with the most frequently confused items.

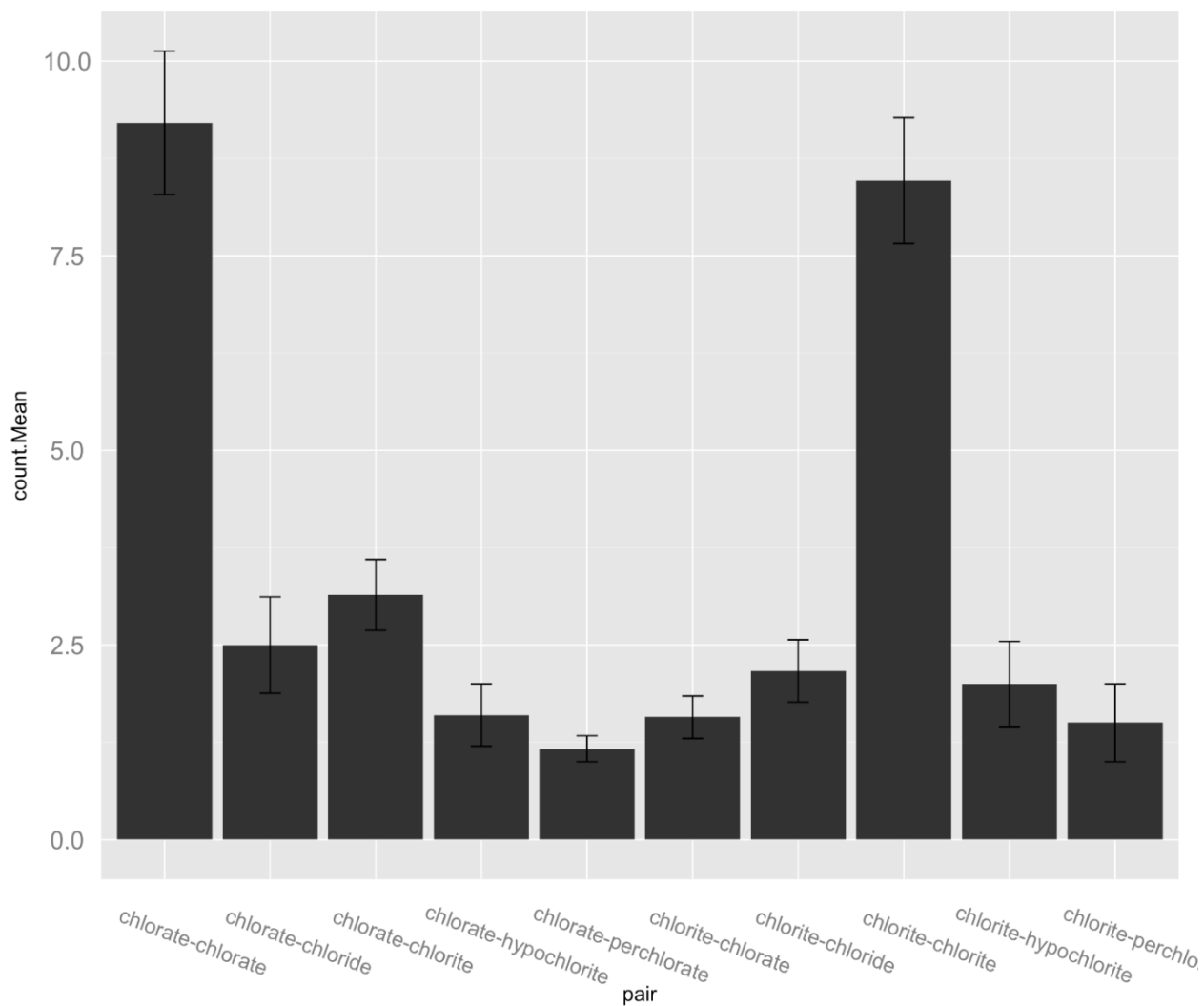


Figure 3. Frequency of confusion for the ions chlorate, chlorite and chloride, averaged across participants.

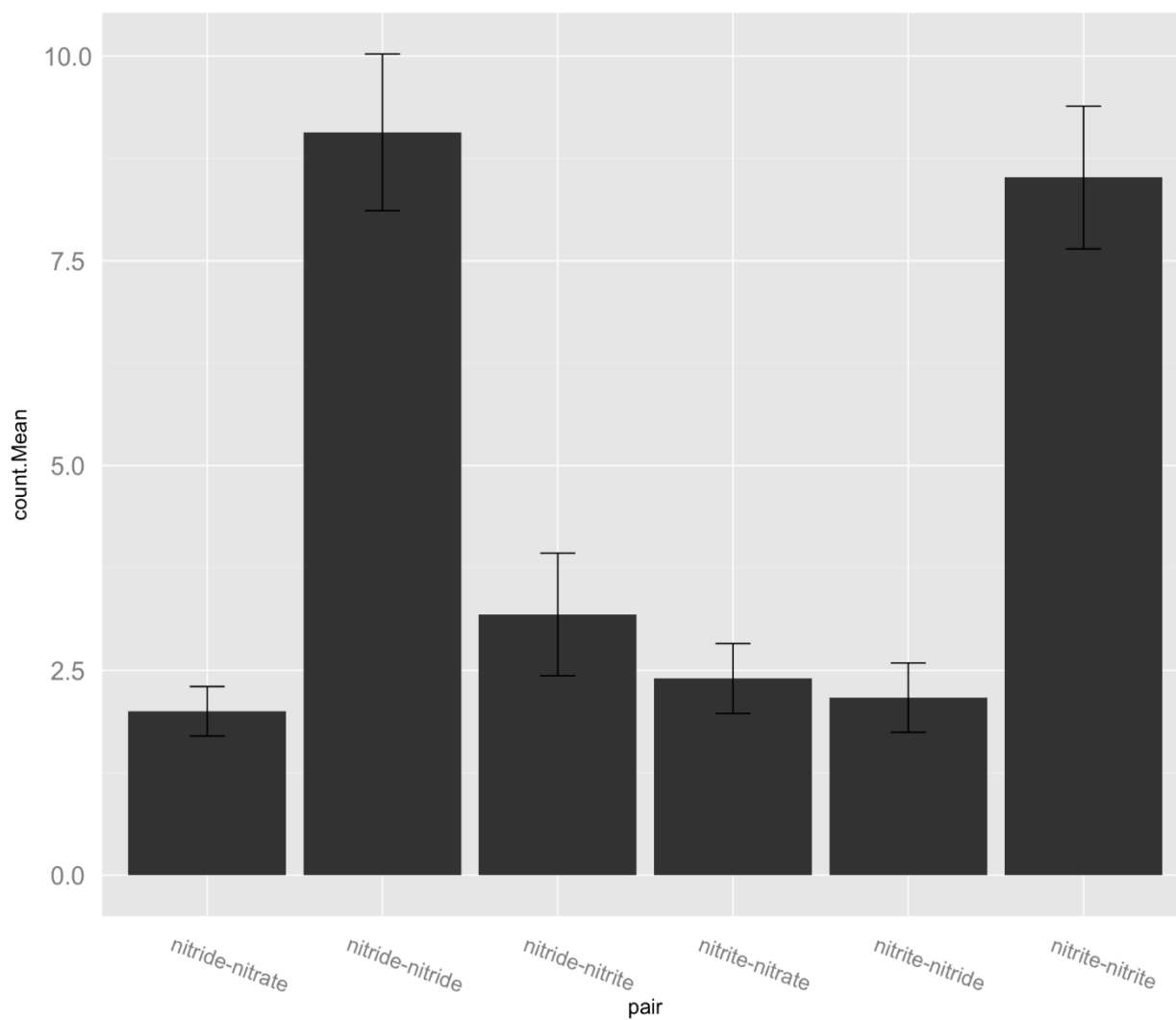


Figure 4. Frequency of confusion for the ions nitrate, nitrite, and nitride, averaged across participants.

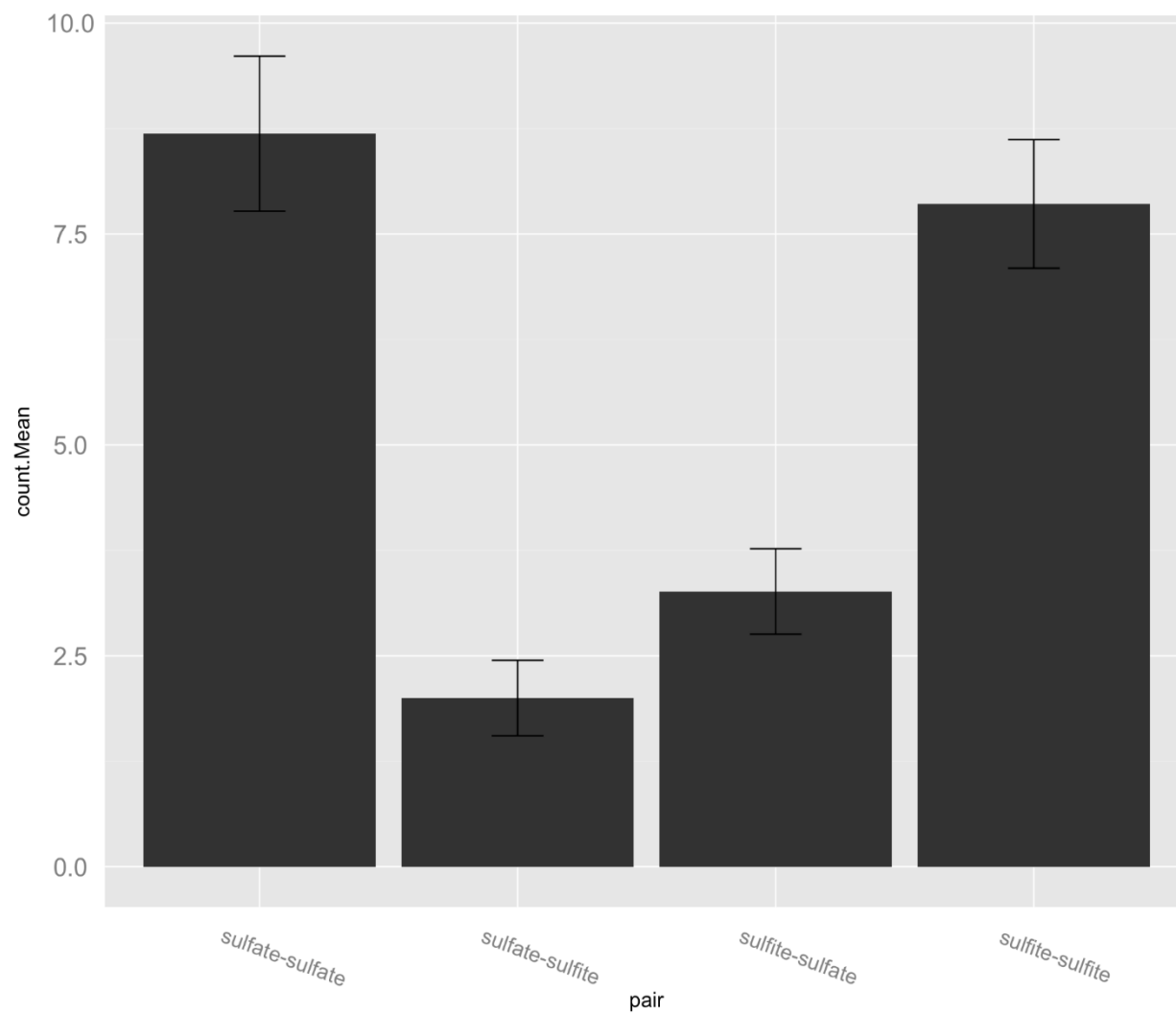


Figure 5. Frequency of confusion for the ions sulfate and sulfite, averaged across participants.

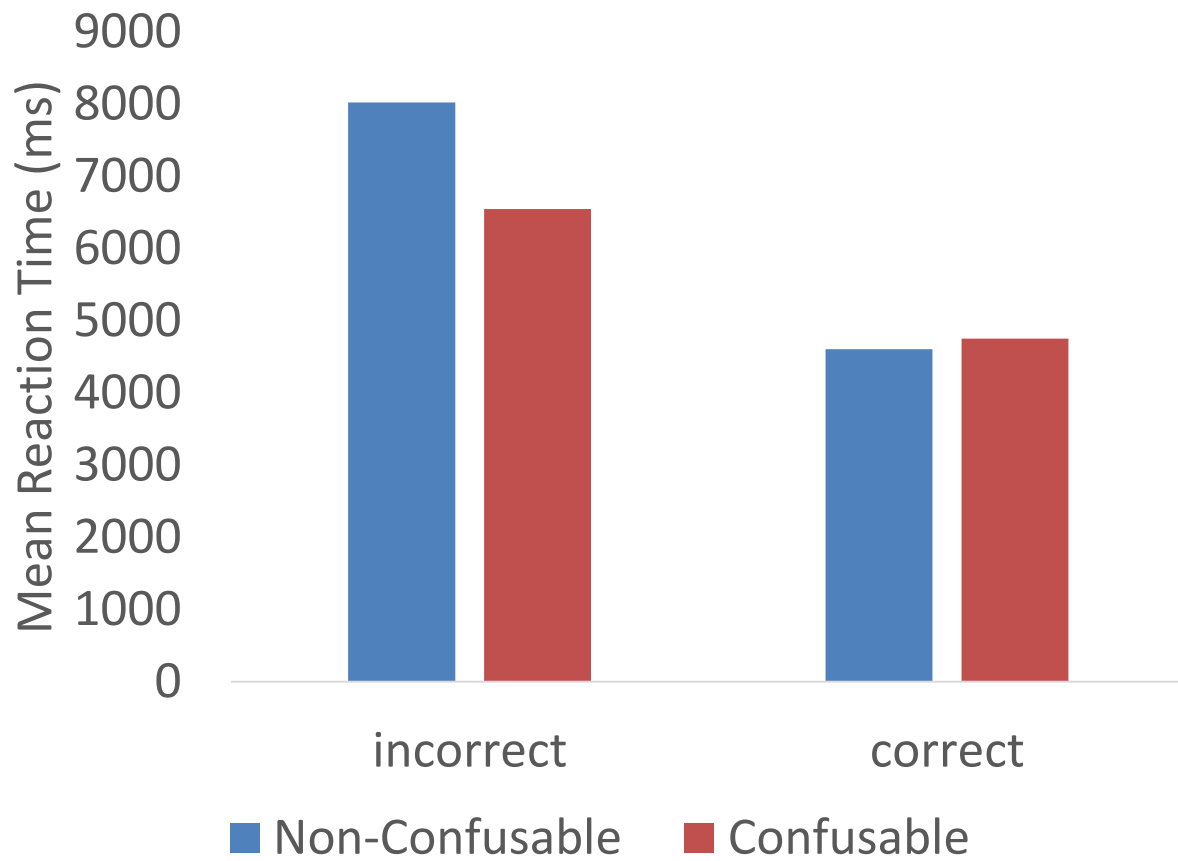


Figure 6. Mean reaction time of incorrect and correct responses for confusable and non-confusable items.

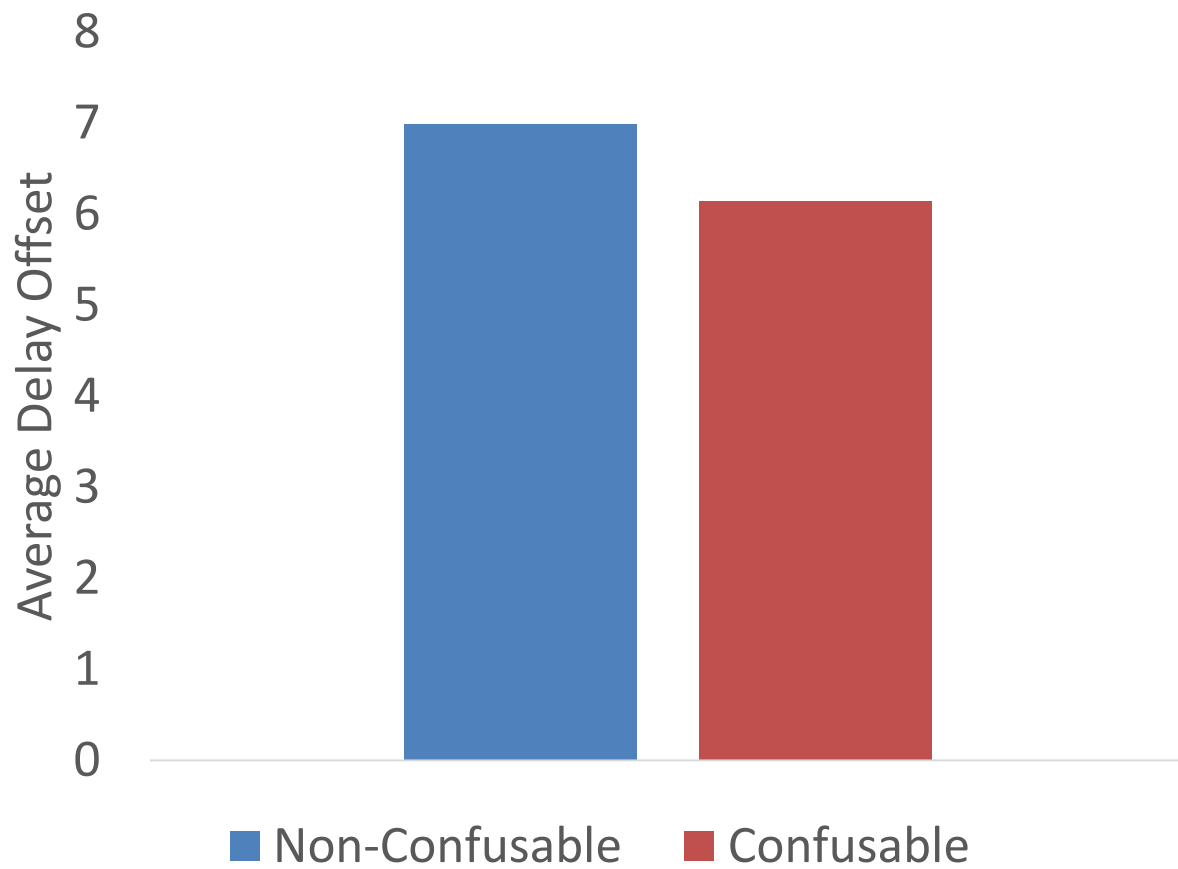


Figure 7. Average delay between repeated tests of confusable and non-confusable items.

Remember-Know Judgments and the Attention to Memory Hypothesis

Kaitlin Swinnerton

Memory retrieval can be divided into two different categories, recollection and familiarity. Recollection is characterized by high confidence and the presence of accompanying memories for episodic details of the moment of encoding. Familiarity lacks the presence of accompanying episodic memories, but can still exhibit high confidence judgments. While it is generally agreed that there is a difference between recognition and familiarity based retrieval, the underlying mechanisms that contribute to these different subjective experiences differ. There have been many studies that look to more clearly characterize this distinction,

One such study was done by Dudukovic & Knowlton (2006). The goal of the experiment was to differentiate between two competing hypothesis that both defined the recollection v familiarity distinction. The first hypothesis was the signal detection theory model that distinguished recollection and familiarity in terms of trace strength. In this model, recollection based memories simply passed a higher confidence threshold than familiarity-based memories. The second hypothesis was the dual process model, which argues that recollection and familiarity are two distinct processes where recollection is characterized by the recall of episodic details and familiarity is uses a signal detection criteria.

In order to distinguish these hypotheses, subjects completed remember-know tasks based on a list of word pairs they studied. Each pair was presented with both line drawings depicting the word and the two words printed at the bottom. One word was a cue word and the other was the target. The cue words varied in their color and orientation on the page. After the study period, subjects first completed a remember-know task after a delay of 10 minutes. The task presented

subjects with either target words from the list they just studied or new words that did not appear on the list. Subjects first had to determine whether the word was old or new. If they answered that it was old, they then had to indicate whether they “remembered” the word, meaning they could recall episodic details about learning the word, or whether they just “knew” that they had seen the word. Remember judgments correspond to recollection and know judgments correspond to familiarity. Subjects returned a week later and completed another remember-know task. At that time, they were also quizzed about the color, orientation, and cue word for each target. These questions were used to determine the amount of episodic details remembered for each target and thus test to see if remember judgments were in fact associated with increased memory for episodic detail.

As expected, the researchers found that remember judgments were accompanied by more memories for episodic detail than know judgments. Additionally, items that received know judgments at one week were not accompanied by episodic memories, even if they had received a remember judgment at the 10 minute delay. This suggests that the transition from remember to know reflects a loss of memory for episodic detail. Furthermore, there was an interaction between delay length and response type, such that the number of remember judgments decreased sharply over the longer delay, whereas the know judgments stayed relatively stable. This interaction suggests that recollection and familiarity are distinct processes. In sum, the results from this study support the dual process hypothesis, that recollection and familiarity are in fact separate processes.

Another study that supports this hypothesis is one done by Ciaramelli, Grady, & Moscovitch (2008). This study reviewed existing fMRI data and developed a model to explain the process of episodic memory retrieval by relating it to attention. They introduce the

component process model, which distinguished between direct and indirect memory retrieval. Direct retrieval is an automatic process in which a cue accesses memories via interaction with the MTL. The process requires no additional attentional resources. However, indirect retrieval is an attention demanding process in which a cue does not automatically elicit the target memory and additional strategic search processes, which are mediated by the prefrontal cortex, must be used to access the target memory.

Additionally, there are two distinct attentional processes that are differentially relevant to memory retrieval. The top-down system uses the superior parietal lobe and the intraparietal sulcus. It is involved with voluntary attentional focus. It uses top-down information, such as the individual's personal goals, to identify which aspects of the environment require attention. The bottom-up system uses the inferior parietal lobe and the temporo-parietal junction. It automatically allocates attention to elements of the environment that are important to the individual.

The authors synthesize these models to create their own Attention to Memory model. They hypothesize that direct retrieval relies on the IPL and the bottom-up system. IPL should then be the most active when subjects are experiencing recollection, high confidence judgments, and generally strong memories. On the other hand, indirect retrieval relies on the SPL and the top-down system. The SPL should be most active when subjects are experiencing familiarity, low confidence judgments, or anything that requires the subject to engage in additional processes to make a memory decision.

The authors review many different types of experiments in order to support their model, but the most relevant to this discussion is the remember-know study discussed. They predicted

that recollection would lead to high confidence judgments and high IPL activity. However, familiarity judgments were predicted to require more effort in distinguishing between old and new because they lack the extra episodic details characterizing recollection. Thus, greater SPL activity would be expected in familiarity judgments because more pre and post retrieval processing would be required to correctly determine if something was familiar or new. They found that the IPL was increasingly active for recollection v familiarity judgments and that the SPL was more active compared to the IPL in familiarity judgments. This data supports their Attention to Memory model. It also provides evidence that recollection and familiarity are two distinct processes characterized by differential brain activity. This provides evidence for the dual process model supported by Dudukovic & Knowlton.

My Experience in the Lab this Quarter

Overall, my experience in the lab this quarter was positive. I especially enjoy the lab meetings, even if I didn't always understand everything that was being discussed. The first two meetings, in which Andrew and Nicco presented papers, were particularly interesting. I liked that the papers were emailed out so that I could review them before the meeting. It allowed me to follow along with the discussion more easily and I feel like I learned a lot from those two meetings.

That being said, like last quarter, I wish we could have had more meetings throughout the quarter. I did appreciate when information about talks happening on campus and such would be sent out, but my hectic schedule prevented me from being able to attend many of them. Hopefully in the future I will have more opportunities to attend these events because I think I could learn a lot from them.

As for my day to day experience running subjects, it was very similar to last quarter. Even if my job itself isn't the most exciting, the studies are all looking at interesting questions, and I'm excited to be part of them. I'm interested to see how what the results will show and what implications they might have.

Even though I am graduating this quarter, I would be interested in staying involved in the lab if possible. I enjoy being involved in the research and am excited to contribute however I can. While I have no concrete plans for grad school yet, it is definitely something I am strongly considering, and my experience in the lab has given me a good perspective of what a future working in research might entail. I feel very lucky to have been a part of this lab the past two quarters and I can't wait to see what new knowledge the lab will produce in the future.

References

- Ciaramelli, E., Grady, C. L., & Moscovitch, M. (2008). Top-down and bottom-up attention to memory: a hypothesis (AtoM) on the role of the posterior parietal cortex in memory retrieval. *Neuropsychologia*, 46(7), 1828-1851.
- Dudukovic, N. M., & Knowlton, B. J. (2006). Remember–know judgments and retrieval of contextual details. *Acta psychologica*, 122(2), 160-173.

Foreign Language Learning Through a Virtual Environment

Katherine Elizabeth Guardado

Whether it be the field of Humanity Communication, or Medicine, Technological advancements have strived to create a better world for society and its inhabitants. Various advancements have been invented to aid individuals with daily tasks, such as learning and work to recreate and integrate technology. Direct application of this concept can be seen in the case of training in virtual environments (VET). A virtual environment is a 3 dimensional deception of an environment and permits individuals to move around and direct themselves as if they were present in the scenario. For example, VET can be used for training of pilots, fire fighters, drivers, control operators, and second language learning (Moskaliuk et al, 2012).

Virtual environments are highly advantageous as they could be used when circumstances of training in the real setting could be dangerous or very costly. Furthermore, they provide high potential for language learning and teaching in settings that go beyond the standard classroom setting. VE provide students with conversational opportunities and cultural experiences that enhance learning in manners that are limitless (Chen, 2011).

A Study conducted by Ya-Chun Shih and Mau Tsuen Yang revolves around having a collaborative virtual environment that promotes language learning. They have created a virtual 3D online communication environment that has been fostered to help undergraduates who are learning English as a foreign language. The 3D environment includes an online stimulator, which allows learners to communicate and immerse in a 3D interactive classroom that resembles reality. This classroom has been named VEC3D and it integrates goal based instructional learning, engaging 3D graphs and real time voice communication with other students in the classroom. The ultimate goal of the program is to determine how learners perceive their

experiences in the virtual space and study whether there is a growth in their understanding of English. Upon conducting the study, the results obtained reveal that the proposed application promotes a positive student attitude and interactive learning experience. Students have been found to understand English better and feel more comfortable communicating with others in their actual classrooms because of the support they receive from this virtual program (Chun & Yang).

For the current study purposes graduate students Joey Essoe and Nicco Reggente developed an avatar learning virtual environment using OpenSimulator. OpenSim is commonly used for research purposes as it can be used independently of the influence of others by having a private server (Chen, 2011). Using the virtual environment called Rissland participants were trained to learn a list of 42 Swahili words. For each participant run the experiment lasted 4 days. From day 1 to day 3 the experiment was performed in a room where the windows were covered with foil paper so light would not penetrate in. This was done in order to create an atmosphere where participants would feel immersed in the virtual environment. The subjects also used a headset and a joystick to guide their avatar. We used Swahili words since most people are less likely to know Swahili compared to other languages such as Spanish or Chinese, hence decreasing maturation in the study.

The VE consisted of the “AVATAR center” which was the main area where participants either males or females could change their avatar and make them look like themselves. Moreover, there were 9 different locations where subjects could move their avatar through a path with a variety of environmental cues, during learning or testing sessions the participant sat their avatar in a specified location and a screen was prompted to them with a presentation of the 42 images with the word written in Swahili and in English and also pronunciation in Swahili which they had to repeat for a couple of times before moving on to the next word. The order in which

the items where given was randomized and different for each participant. How well participants remembered the Swahili words was determined by a scoring system in which each syllable in a word was equivalent to a point, if the order of the syllables was changed then .25 points were subtracted, if the vowel in a syllable was remembered but not the consonant than .5 points were subtracted and vice versa when the consonant was remembered but not the vowel.

On day one participants had three learning sessions and three testing sessions. There were 2 different conditions: varied and fixed. In the varied condition participants were allowed to explore a determined location, and then they were given the learning session in the same location. Then, they explored a different location and were tested in that location. This process was repeated for a number of three times on day one (Table1). In the fixed location, participants were given a specific exploration location and then they were given the testing session in a different location. Then, they explored another location and went back to the same location where the learning session was administered.

On day two participants had two testing sessions and one learning session. The varied condition was the same as the above varied condition and the fixed condition was the same as above as well. On day 3 participants were taken to a different room, one of the experimenter's office as we thought this would be a neutral environment. The experimenter asked subjects to recall the 42 Swahili words they preciously learned. Then, day 4 was performed 18 days after day 1 on a bench in front of Franz Hall at UCLA. There, a second experimenter tested the participants on the 42 Swahili words learned on days 1 and 2. Recently, a day 5 has been added to our study, which is done 100 days after day 1. On day 5, a third experimenter calls participants. In order to avoid practicing effects they are not aware they will be tested on the 42 Swahili words they learned, but they are told the call will be just a survey about their experience

as participants of our study. This new addition to the experiment will help us understand the effect of virtual environment learning on long-term memory.

Even though I was not part of the data analysis which is currently being done by other lab members. From the data I collected it could be correlated that a virtual environment does enhance memory as by day 4 the participants remember most of the Swahili words they learned. Moreover, I was able to observe that in order for the virtual environment training to be effective, people using it need to feel identified with it. They need to feel comfortable enough otherwise using an avatar (a representation of themselves); otherwise, the training might not work. This was the case of one of the participants who I ran who was clearly not interested in learning through a virtual environment as a result he learned less than 5 words by the end of the training compared to the rest of the participants

By using pictures for the learning sessions we investigated the effect of the usage of pictures to learn a new language. According to Carpenter et al it has been long believed there is a picture superiority effect which states that pictures are remembered better than words given that it is easier to relate a word to a picture than to form a mental image of the word. However, some disagree arguing pictures do not give an advantage over native language translations of words into another language. They performed a similar study where they tested participants on 42 Swahili words learned through pictures presented to subjects on a computer. Swahili words were learned better from pictures than from translations. They concluded that pictures facilitates learning of foreign language vocabulary as long as participants are not too overconfident in the power of a picture to help them learn a new word (Carpenter, 2011).

While various studies have been conducted on learning and their relation to language, our study aims to understand the effect that personalized virtual environments have on learning and

how it may contribute to the success of individuals learning new languages or even learning new topics that they are not familiar with. With the results and data of our study, we might be able to introduce a new outlook on learning and create a correlation between learning and virtual environments. With this advancement, individuals who are learning different languages or concepts have the opportunity to learn in a manner that is quite different than traditional way, which is emphasized presently in society. Contrary to the traditional manner, this virtual setting allows individuals to go beyond the classroom setting and captivate in the process of learning which is more personally built and thus, better assess their needs. With technological advancements, the approach that individuals have towards learning are quite different than prior to technology and with this change in approach, virtual environments might provide a novel method to promote learning as it works to integrate technology and the daily tasks that individuals perform.

Figure 1



Table 1. Varied Condition

	Change BEFORE	Explore BEFORE	Test/Learn
L1	Rusty (Animal)	Town	Town
T1	Dexter (Male)	Swamp	Swamp
L2	Dragon (Animal)	SpaceGarden	SpaceGarden
T2	Miss Smith (Female)	BoraBora	BoraBora
L3	Wolf (Animal)	UnderSea	UnderSea
T3	Dr. Chase (Male)	BigBear	BigBear
T4	Bumble Bee (Female)	Garden	Garden
L4	Andrea (Female)	Snow	Snow
T5	Prince (Male)	Country	Country

Table 2. Fixed Condition

	Change BEFORE	Explore BEFORE	Test
L1	Rusty (Animal)	Town	Town
T1	Rusty (Animal)	Swamp	Town
L2	Rusty (Animal)	SpaceGarden	Town
T2	Rusty (Animal)	BoraBora	Town
L3	Rusty (Animal)	UnderSea	Town
T3	Rusty (Animal)	BigBear	Town
T4	Rusty (Animal)	Garden	Town
L4	Rusty (Animal)	Snow	Town
T5	Prince (Male)	Country	Country

References

Carpenter C, Olson K. (2011). Are Pictures Good for Learning New Vocabulary in a Foreign

Language? Only If You Think They Are Not. *Journal of Experimental Psychology*.

38(1), 92-101.

Moskaliuk, J., Bertram, J., & Cress, U. (2002). Training in virtual environments: putting theory

into practice. *Ergonomics*, 1-10.

Moskaliuk J., Bertram J, Cress U. (2013). Impact of virtual training environments on the

acquisition and transfer of knowledge. *Cyberpsychology Behavior and Social*

Networking. 210-214.

Shih, Y., & Yang, M. A Collaborative Virtual Environment for Situated Language Learning

Using VEC3D. *Educational Technology & Society*, 2(1), 56-68.

Videogames & Motor Learning

Kylie Springsteen

Videogames have become an increasingly prominent part of our culture in the past few decades. Many games are violent, particularly including shooting violence involving guns. They are becoming more violent as computer graphics have advanced and become more realistic, and many games actually implement pistol-shaped controllers. Whitaker & Bushman (2012) explored the effects of violent videogames, specifically, the effect of video game play and controller type on firing aim and accuracy. That is, can violent shooting games be used to train people to more accurately and lethally fire a realistic gun?

Videogames utilize many features known to promote learning including motivation, instant feedback, reinforcing behaviors, and adapting to skill level (Gentile & Gentile, 2008). Additionally, the element of interaction is a key feature of videogames that enhances learning. Research shows that learning is increased when the learner is actively involved in the task (Onion & Bartzokas, 1998). Compared with watching a movie, where the viewer passively observes others acting aggressively, in videogames, players themselves actively become aggressive characters. Additionally, research shows that aggressive actions increase subsequent aggressive actions (Bushman, 2002). Videogames are used specifically for the purpose of teaching in some professions, pilots and surgeons, for example. In fact, the U.S. military even uses videogames to train soldiers. This indicates that videogames can be used as successful teaching tools and begs further explanation of their effects. No previous research has been done to explore the transfer of skills, which seems to occur through the playing of videogames.

In this study, participants were randomly assigned to one of five conditions. There were three types of games varying in their amount of violence. The first was a violent shooting game, Resident Evil 4; this game used humanoid targets and also rewarded headshots. The second was a nonviolent shooting game, a Wii Target Practice game; the participant shot at Bull's Eyes. The third game was a nonviolent nonshooting game, Super Mario Galaxy; no shooting was involved. For each of the shooting games (Resident Evil 4 and Wii Target Practice) participants were randomly assigned to use either a pistol-shaped controller or a standard controller. For the nonshooting game, participants used only the standard controller. Each participant played his assigned game with the assigned controller for a twenty-minute section of the game chosen specifically for its amount of continuous shooting, approximately 300 shots in the twenty minutes.

In the posttest, each participant shot sixteen bullets at a 6-foot mannequin from 20 feet away. A black airsoft training pistol resembling a 9mm semiautomatic was used. It is important to note that the posttest only design was used to eliminate practice effects. Accuracy of shots as well as the number of headshots taken were measured as an indication of learning. After controlling for aggression traits, attitude toward guns, firearm training, and previous exposure to shooting games, Whitaker and Bushman found that the people who played the violent shooting game using the pistol-shaped controller had significantly higher accuracy and higher number of headshots than the people who played the nonviolent nonshooting game with the standard controller. These results indicate that there is some transfer of motor skills between the videogame task and the shooting task. Additionally, a positive correlation was shown between the amount of previous exposure to shooting games and the accuracy of shots taken. This indicates the possibility of a long-term effect in that more practice on the videogame, produces

better real life shooting. However, correlation does not indicate causation and further experimental exploration must be conducted in order to make conclusions about this phenomenon.

One major confound in this study is that it used a violent shooting game that rewarded headshots. In this case, the significant difference in number of headshots cannot be attributed to the amount of violence or the use of humanoid targets in the game. It can only be attributed to operant conditioning, as the player learns to be more accurate in taking headshots due to the use of positive reinforcement for doing so. It can be supposed that if the game used did not reward headshots, the difference in the accuracy of headshots may not be as pronounced.

Another problem in this study is that their sample consisted of 51% males. College undergraduate males spend a lot more time, on average, playing video games than their female counterparts, perhaps producing skewed results. Although the groups were controlled for previous amounts of exposure to shooting games, the results might be more reliable if the researchers included a sample of, for example, all women who have never played video games and showed that playing the shooting game improved accuracy of shooting.

Overall, this study does not show that people are more violent or more likely to fire after playing the game. The only implications of these results are that if a participant were to fire, he would be more accurate and more likely to aim at the head after playing a violent shooting game that rewards headshots. Instead, this study serves as an exploration of motor skill transfer, not of violence. Further, it is conceivable that any type of videogame could improve hand eye coordination and accuracy of firing from pretest to posttest.

The current posttest design is limited in its implications. As is, it has only shown that the condition with the most similar elements to the posttest task led to greatest transfer of motor skills. The game that used humanoid targets and a gun controller had greatest transfer to the posttest task of shooting a gun at a humanoid target which does not imply effects of violence, rather only shows a transfer of motor skills. Another type of posttest using a different target, or a different shooting mechanism should be added to this study. Given a posttest requiring the player to shoot darts from a gun at a Bull's Eye, the group who trained on the Wii game with the pistol controller could have the highest accuracy due to the greatest amount of similar elements between tasks. Alternatively the posttest could use a completely different motor task without similar elements to determine the true effect of each condition on motor skills. Additionally, if a reward was added for headshots in the posttest, the other conditions shooting more headshots might be observed. The Resident Evil group had already been conditioned for headshots but if the difference between groups decreased in significance when rewarded for headshots, nothing could be attributed to the particular features of the game used. This would help to determine if the effects were indeed operant conditioning, or if the violence somehow affects individuals. Although several posttest alternatives might be explored, the current posttest design only shows that there is greater transfer for more similar elements and does nothing to explain the effect of any of the other game components.

Zililab also studies motor learning. Chela's experiment explores the effects of different practice schedules on motor learning. The varied group practices throwing at 5 and 9 feet, while the specific group practices throwing only at 7 feet. Then, all participants are tested throwing at 3, 5, 7, 9 and 11 feet. Whitaker and Bushman (2012) showed in their experiment that the condition with the highest number of similar elements leads to greatest task transfer. In Chela's

experiment, it is less of a matter of task transfer, because the posttest task is the same task as the practice task. However, by this theory of similar elements, it could be predicted that the varied group might have higher performance because there are more similar elements between their practice schedule and the posttest. The varied group has previous practice throwing at two of the five distances while the specific group has previous practice throwing at only one of the five distances; therefore, it is predicted that the varied group will show higher performance. This is similar to the videogame experiment in that the Resident Evil group had previous practice shooting at humanoid targets while the Super Mario Galaxy did not so they did better on a task where they were shooting at a humanoid target. Overall, the connection between these two experiments is that previous practice in a task leads to better performance. However, Chela is not finding such commonplace results and this has different implications for motor learning than the findings of Whitaker and Bushman (2012).

References

- Bushman, B. J. (2002). Does venting anger feed or extinguish the flame? Catharsis, rumination, distraction, anger, and aggressive responding. *Personality and Social Psychology Bulletin*, 28, 724-731.
- Whitaker, J. L., & Bushman, B. J. (2012). " Boom, Headshot!": Effect of Video Game Play and Controller Type on Firing Aim and Accuracy. *Communication Research*, 0093650212446622.
- Gentile, D. A., & Gentile, J. R. (2008). Violent video games as exemplary teachers: A conceptual analysis. *Journal of Youth and Adolescence*, 37, 127-141.
- Onion, C. W. R., & Bartzokas, C. A. (1998). Changing attitudes to infection management in primary care: A controlled trial of active versus passive guideline implementation strategies. *Family Practice*, 15, 99-104.

IMPORTANT NOTICE: The following is an undergraduate research paper submitted as part of the requirements for one quarter of participation in a research apprenticeship in the UCLA Human Perception Laboratory (HPL) in the Department of Psychology at UCLA. This document has not undergone a scientific review and may or may not relate to or correctly represent projects currently or previously undertaken in the HPL. Some features of adaptive and perceptual learning techniques mentioned or described herein are protected by US Patent #7052277 and patents pending. All rights reserved. For information, contact info@insightlearningtech.com.

Effects of Active and Passive Presentation Style on Factual Learning

Lorena Hinckley

Historically, text has been the dominant format for teaching factual material and books have been the primary teaching 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 benefit of presenting information in an interactive way that engages the learner instead of a traditional passive presentation style.

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 and 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 et al. (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 two possible outcomes of learning from an interactive learning module: the active-learning hypothesis and the passive-learning hypothesis. The passive-learning hypothesis predicts that learning in an interactive environment has no effect on learning because it is the informational content that matters, and the content is the same in an interactive setting as in a non-interactive setting. Alternatively, the active-learning hypothesis 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 and Wolvin, 1997). As a result, the active-learning hypothesis predicts that learning should improve when students use interactive learning programs.

In contrast, the passive-learning hypothesis predicts that the interactivity of a learning program should not affect the level of learning. This hypothesis is derived from the transfer model of learning, often used to contrast constructivist model (see e.g. Mayer, 1999). The transfer model argues the learning process involves only the transfer of knowledge from experts (e.g. textbooks) to the learner with the learner's role being only that of a passive recipient (Mayer, 1999; Evans and Gibbons, 2007). The delivery method of the information is irrelevant; the only thing that matters is the quality of the informational content, thus the level of learning should not differ between interactive and non-interactive environments.

An Experiment

In this experiment, participants study a set of 24 African countries. For one group of participants the countries are presented in an active format, but for the other group the study

session is set up as a passive learning task. Participants in the passive group study each country for a short period of time during which the correct country name is also displayed on the screen. Both groups of participants complete the same geography test before the learning phase, immediately after completing the learning phase, and then again one week later.

Method

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 a list of the names of the 24 countries in alphabetical order.

Design and procedure This experiment has two conditions manipulated through which module the subjects are given: active and passive. The effect of these different modules on the performance of participants on the post tests is measured in terms of accuracy and response time, measuring recall ability and fluency of the material. Both of the modules consists of a pre-test, a training phase, a post-test, and a delayed post-test.

The pre-test, post-test, and delayed post-test are all identical in format. They each consist of 24 randomly ordered questions (one per country). For each question, a map of Africa with one country highlighted is displayed on the screen, and the participants are asked to select the correct name of the highlighted country from the list of countries on the right. Participants are not provided with feedback. The pre-test is given immediately before the training phase, the post-test is given immediately after, and the delayed post-test is given one week later.

During the training phase, participants complete either a passive or an active learning module. For the active learning group, participants are presented with a stimulus and are instructed to select the correct name of the highlighted country. Immediately after choosing a

country name, the participant is given feedback. The stimuli are blocked into groups of ten, and are presented in a random order determined by an algorithm that adapts based on the participants performance (Mettler et al., 2011). Participants are trained until they reach a certain learning criterion based on accuracy and response time. For the passive group, participants are presented stimuli with the correct name for the country highlighted in the list of countries and written above the country on the map. Stimuli are presented for five seconds before disappearing from the screen, after which participants are prompted to continue. Stimuli are blocked into groups of ten, and the number of blocks a participant sees is yoked to the number of blocks a participant in the active condition uses to complete the module.

General Discussion

In this experiment, the active-learning hypothesis predicts that the participants in the active condition will outperform those in the passive condition on the post-tests because of the effects interactivity has on learning. Alternatively, the passive-learning hypothesis predicts there will be no significant difference in the test scores of the participants in the passive group as compared to the participants in the active group because the information presented in the two conditions is the same. The delivery method and interactivity do not matter. Based on the prior research discussed earlier, the active-learning hypothesis is predicted to be supported in this experiment.

The effects of interactivity on computer-based learning has important implications for the design of future programs. Currently, many e-learning systems are non-interactive 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, incorporating more interactive elements into their systems.

References

- Coakley, C., & Wolvin, A. (1997). Listening in the educational environment. *Listening in everyday life: A personal and professional approach*, 2, 179-212.
- Evans, C., & Gibbons, N. J. (2007). The interactivity effect in multimedia learning. *Computers & Education*, 49(4), 1147-1160.
- Evans, C., & Sabry, K. (2003). Evaluation of the interactivity of web-based learning systems: principles and process. *Innovations in Education and Teaching International*, 40(1), 89-99.
- Mayer, R. E. (1992). Cognition and instruction: On their historic meeting within educational psychology. *Journal of Educational Psychology*, 84, 405-412.
- Mayer, R. E. (1999). Designing instruction for constructivist learning. *Instructional-design theories and models: A new paradigm of instructional theory*, 2, 141-159.
- Mettler, E., Massey, C., & Kellman, P. J. (2011). Improving adaptive learning technology through the use of response times. In *Proceedings of the 33 rd Annual Conference of the Cognitive Science Society*. Boston, MA: Cognitive Science Society, 2532-7.
- Moore, M. G. (1989). Editorial: Three types of interaction. *The American Journal of Distance Education*, 3, 1-6.
- Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents?. *Cognition and Instruction*, 19(2), 177-213.

Confidence Weighting on Multiple Choice Tests

Monica Vu

The acquisition and consolidation of learned material is a topic that concerns many students and teachers alike. There are often misconceptions and clashes in what is believed to be the most efficient way to study and test information. The Bjork Learning and Memory Lab examine the different mechanisms in hopes of finding optimal strategies to learn. For example, when students first learn a topic, they tend to read and re-read that material. While many would find this process helpful, studies have shown that the familiarity of the topic becomes low-level priming (Bjork 2011). Many students undergo an illusion of comprehension when the material appears to be easier. This produces an example of what is called ‘retrieval strength’ that could be reflected in test scores following examinations. It is storage strength and long-term recall however that determines the real consolidation of memory. Learning is considered to only be acquired when there is some level of difficulty, a process coined by the term that is often referred to as “desirable difficulty.” There are a number of ways to implement this desirable difficulty which could include spacing, varying conditions, and the topic that we’re most concerned with, testing.

Spacing is a mechanism that has been well-studied and established means for consolidating material (Kornell & Bjork 2008). Spacing is learning material over a long period of time along instead of rapidly trying to cram. Interleaving on the other hand is a tactic that some might find counterintuitive. Interleaving is done by studying material in a mixed up order instead of in ‘chunks’. When learning new material, most would prefer to study one subject extensively before moving onto the next. Novel conditions however proved to be much more effective. This could be due to having a holistic view on things and making multiple connections

to what is being learned. Most people who prefer the chunking method may feel that similarities are easy to detect. The ease and familiarity now of the subject is typically attributed to learning but usually it is not the case. The study by Bjork and Kornell tested their participants on paintings of certain artists through a memory recall test. The participants either studied paintings that were grouped by the artists or in an order that puts the artists in between each other. Even though the participants felt that they did better with a chunking condition, in actuality they performed better when the material was presented in an interleaved manner.

A process used by almost all instructors is testing. A debatable form of testing is a multiple choice format. This topic is ubiquitous considering how many assessments are done through this format. Some, if not all, portions of standardized testing, SAT's, LSAT's, MCAT's, and many other exams use this format. A few of the main reasons why this form of testing is used so often mainly is due to the convenience obtaining scores and how quick it is. There are many disagreements with this format, claiming that there is no engagement with the material or that it does not support long-term retention. Little et al. (2012) is a recent article that tries to debunk the harsh criticisms that some educators have against multiple choice testing. The article shows that if given incorrect choices that are plausible, the multiple choice test could actually be helpful in assessing whether they do or do not know the information well. Our study also hopes to find that multiple choice testing can be very effective.

Modified Confidence Weighted – Admissible Probability Measurement, or MCW-APM, is an alternative approach to multiple-choice testing (Bruno 1989). In this study, the students in an economic class were tested on their confidence of learned material. The format of the questions and answers are very similar to the method we use in our experiment. The answer choices are given in the form of a triangle with a finite number of options. Three options on each

point are considered the extreme answers with the correct answer having the highest amount of points and the other two and all options between them having the lowest score. In between options would have a varied number of points based on position. “Near misinformed” would be answers that are closer to the correct one, “misinformed” are on the opposite side of the correct answer, and “uninformed” is an option that is isolated from all the others and in the direct center of the triangle. Results showed that students seldom picked extreme answers despite given more points. It was also evident that these types of exams are more effective than just a simple right-wrong approach. This is believed to be because the students are thinking more critically about the material and ultimately understand it better.

Given the evidence that multiple choice testing can be effective if done correctly, we wondered if there was a benefit to adding a confidence-weighting answering process. Our hypothesis is that students will do better given pre-tests than if they were to only study a material before being tested on it and that a confidence-weighted would produce better results than a regular multiple-choice test.

Method

Participants

When first reported at the end of last quarter, only 69 students had participated in our experiment. This number was insufficient to provide us data that was statistically sufficient. Spanning into this quarter, there was a total number of 150 participants in this experiment. All of the participants were students from the University of California, Los Angeles and were only given class credit for their participation. These students were recruited using the online SONA system.

Design and Materials

This experiment was a one-way between subjects design. The independent variable that was varied through our experiment was the format in which the subjects received the material to learn. There were three available conditions. The participant either had standard-multiple choice, confidence- weighted multiple choice, or study only. In the standard multiple choice, the student was given a multiple choice test prior to answer questions. The confidence weighted-multiple choice option was the same as the regular multiple choice but instead the participants were tested using a triangle similar to the MCW-AWP format. The confidence weighted-multiple choice pre-test had a total of 13 possible options to choose while the regular multiple-choice gave three options. The study only condition had no-pretest. This experiment was done through a webpage online. All of the participants were given the same two-articles to study and be tested on, one on Yellowstone National Park and one on Saturn.

Procedure

The participants did the experiment individually on separate computers. A web link that was unique to the experiment was used. Demographical questions were asked followed by instructions on what would occur and how to do the study. Participants in the confidence-weighted multiple choice condition were given the MCW-APM triangle to become acquainted with. They were shown how certain answers corresponded to specific scores. Pre-tests were done to ensure comprehension of the triangle and subjects were only able to continue if they were able to answer all the questions correctly.

The first article was presented (either Yellowstone or Saturn) for 9 minutes. The subjects were encouraged to go back to the article and continue studying for the rest of their allotted time

if they had finished reading early. Following that were either the confidence weighted multiple choice test, regular multiple choice test, or Tetris for their respective conditions. They were given 25 seconds to answer per question. The next article was presented and followed by the same testing. Following that, all of the conditions included a Tetris distractor that lasted for 5 minutes. Finally, the all subjects had a cued-recall test that was fill in the blank along with being self-paced. The subjects were asked to give feedback on the experiment by providing any commentary on their strategies and overall opinion.

Results / Discussion

The final score for the study only condition was 4.92 (range 0-13) or 24.5% (standard deviation = 2.813). The final test of the standard multiple choice condition was 7.02 (range 2-15) or 35.1% (standard deviation = 2.592). Lastly, the confidence-weighted subjects scored 8.4 on the final test (range 4-14) or 42% (standard deviation = 2.449). Relatively high scores on the initial tests corresponded high scores on the final test. ($r = .568$, $t(58) = 4.755$, $p < .001$ for the standard multiple-choice condition; $r = .511$, $t(48) = 4.118$, $p < .001$ for the confidence-weighted multiple-choice condition).

A one-way analysis of variance (ANOVA) exhibits the effect of the initial activity(Tetris, standard multiple-choice test, or confidence-weighted multiple-choice test) on final test performance, $F(2,47) = 20.591$, $p < .001$, $\eta^2 = .219$. Planned comparison t-tests between the standard multiple-choice condition and the study-only condition $t(98) = 3.675$, $p < .001$; $d = .74$ and the confidence-weighted multiple-choice condition $t(98) = 6.631$, $p < .001$; $d = 1.329$ demonstrate that effects from both forms of multiple choice testing versus no test align with prior studies. A planned comparison t-test between the two multiple-choice conditions showed a

significance in accuracy for the confidence weighted multiple choice over the standard multiple choice $t(98) = 2.544$, $p = .013$, $d = .511$.

Another factor that we measured was questions that were answered correctly on the final test versus the number of highly confident errors they made. With $r = .486$, $p < .001$, $r\text{-squared} = .237$, the students who indicated highly confident errors tended to score lower on the final test. We think this is due to poor engagement with the material and that little effort was put into retrieval of the information. These particular individuals may not have received the same benefits of the confidence weighted format.

This experiment had some limitations and room for improvement. One change could be to add more passages and questions to acquire a better assessment of the subject's ability. Other directions could be to provide feedback in the pre-tests. We could also look at the effects of using the confidence weighting as a priming tool for standard multiple choice testing. Another possibility could be to look at how this could affect learning conceptual material instead of just facts.

Implications of this experiment could be widely used. Professors and teachers could change the format of standard-multiple choice testing to this confidence weighted design to promote critical thinking of the material while reaping the benefits of a multiple choice test. This method would be beneficial to both educators and students alike.

References

- Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the real world: Essays illustrating fundamental contributions to society*, 56-64.
- Bruno, J. E. (1989). Using MCW-APM test scoring to evaluate economics curricula. *Journal of Economic Education*, 5-22.
- Kornell, N., & Bjork, R. A. (2008). Learning Concepts and Categories Is Spacing the “Enemy of Induction”? *Psychological Science*, 19(6), 585-592.
- Little, J. L., Bjork, E. L., Bjork, R. A., & Angello, G. (2012). Multiple-Choice Tests Exonerated, at Least of Some Charges Fostering Test-Induced Learning and Avoiding Test-Induced Forgetting. *Psychological science*, 23(11), 1337-1344.

Causal Parsimony and the Appreciation of Fractal Artwork

Priyanka Mehta

This quarter, we expanded on a pilot study for an experiment exploring the relationship between the usage of parsimony in making causal decisions and rating the aesthetical appeal of fractal artwork. Our study was based on the central hypothesis that appreciation of visual art in human society evolved at least in part through an appreciation for a general level of complexity that can be applied to both causal and visual appreciation decision making. That is, we predicted that the criteria we use to create causal explanations for events might be related to the criteria we use to decide whether we find something visually appealing. Specifically, the criterion we focused on was parsimony.

In causal reasoning, parsimony is the idea of striving for the simplest explanation that satisfies the most requirements. Previous research has already indicated that we have evolved to make casual decisions not based on the mere fact of covariation of events, but based on actual causal reasoning—an attempt to arrive at a logically sound solution that accounts for all factors present in the event (Liljeholm & Cheng, 2007). But because there is an infinite number of possible causal explanations for any one event, it is important for us to be able to cut out all the far-fetched possibilities and focus in on the most efficient one we can find—and indeed, we do this when we make causal decisions (Liljeholm & Cheng, 2007).

According to Hawking and Mlodinow (2010), parsimony is what drives us to create and improve scientific explanations. Their example illustrating this concept was that of the geocentric-versus-heliocentric theory of the solar system. Hawking and Mlodinow (2010) pointed out that most people believe the heliocentric theory (the sun is the center of the solar system) because they believe it is true, while the geocentric theory (earth is the center of the solar

system) is just plain false, because of countless pieces of evidence in mathematics, physics, and astronomy. However, this is not the case—that is, neither theory is actually “false.” This is because both theories *can* be explained through mathematics and physics and astronomy; however, the heliocentric theory makes these explanations considerably simpler (Hawking & Mlodinow, 2010). While explaining the geocentric theory scientifically would be extremely complicated, the heliocentric theory explains our observations just as well, with fewer rules (Hawking & Mlodinow, 2010). This is the essence of parsimony—people select between explanations of equal validity based on their efficiency and simplicity. It is important to note, however, that an explanation can be simple without being parsimonious, and an explanation can satisfy the most requirements and not be parsimonious. Clearly then, although the idea of parsimony is quite specific, one can see a lot of room for subjectivity in deciding what exactly “simpler” means, and what balance between simplicity and thoroughness is the greatest level of efficiency. The way a question is written can have a large impact on how the participant views his task (Liljeholm & Cheng, 2007), so it is important for the future to make sure our experiment does not allow for ambiguity in places where we do not want it. We encountered some difficulties with this later on during the experiment.

Our proposed relationship between parsimony and fractal artwork also deals with the idea of people choosing options with a preferential level of parsimony. While in causal reasoning, we expect people to choose the most parsimonious explanation to explain something, in art, we expect people to prefer artwork with a pleasing level of parsimony. But how does one measure parsimony in artwork? The way we chose to define this idea is through fractals. Fractals are patterns characterized by “self-similarity”—the idea that different sections or levels of the pattern are highly similar to other sections or levels of the same pattern (Taylor, 2002). For

example, a fractal pattern of triangles might begin with one triangle with three more triangles inscribed in that triangle, and three triangles inscribed in each of those triangles, and so on. Fractals are not limited to artificial geometric patterns, though (Taylor, 2002). Fractals are actually very common in nature. An example of a fractal in nature might be a certain tree, whose large branching structure from the trunk follows a pattern which is more or less repeated in the branching pattern of small branches from each of the large branches, which might be seen again in the branching pattern of twigs from the small branches (Taylor, 2002).

There are two relevant facts about fractals to this study: The first is that they represent order in chaos. This is important because visual art, like paintings, are in a similar situation. Usually (unless we are dealing with geometric art) paintings do not follow geometric formulas; they represent organic forms that do not have a clear order. But, as discussed earlier, natural forms do often contain an underlying fractal order whether we notice it or not (Taylor, 2002). One of our goals in this study was to explore how fractal order in art contributes to how much we appreciate it. But, to appropriately measure such a construct, we needed an art form that was not obviously fractal, but at the same time able to be quantified in fractal terms, which is quite difficult to do with say, the Mona Lisa.

For this reason, we looked at the work of abstract artist Jackson Pollock. Pollock was famous for drip paintings—he would lay out a canvas on the ground and splatter paint all over it in a seemingly random fashion. Although this sounds completely visually unappealing, people were inexplicably drawn towards his artwork, agreeing that there was something aesthetically pleasing about it (Taylor, 2002). Then, a possible explanation for this visual phenomenon appeared: Taylor (2002) found that Pollock's paintings actually contained fractal patterns. In this

way, fractals were linked to modern artwork, bringing us one step closer to understanding why chaotic visual art might be aesthetically pleasing.

The second important fact about fractals is that they have a quality called dimensionality (Taylor, 2002). A fractal's dimensionality has to do with the amount of repeating patterns present in the particular fractal. It is calculated mathematically and is expressed as decimal values between 1 and 2. The greater dimensionality, the more levels of repeating pattern, and the higher the value. So, a pattern with an intricate, dense construction of multiple iterations might have a dimensionality of 1.8, while a simpler, less dense construction of a few self-similar levels might have a dimensionality of 1.2. Because dimensionality is related to complexity, one could say that a low dimensionality is comparable to a high level of simplicity, or parsimony. A high dimensionality is comparable to high complexity, or low parsimony. In this way, we are linking the concepts of dimensionality and parsimony.

We put these facts about fractals and the idea of parsimony in causal reasoning to hypothesize that there is a positive correlation between the way in which we reason causally and the way in which we judge artwork. This means that we predicted that not only would people distinguish between the more parsimonious possible explanation for an event and a less parsimonious one, but that they would also employ the same process in making decisions about how appealing an artwork is. That is, if they preferred extremely parsimonious explanations, they would also prefer parsimonious artworks.

Another important topic we added on to our background for this quarter was the idea of causal invariance. Causal invariance is the idea that something's causal power can be extended across contexts (McGillivray & Cheng). If something is causally invariant, the fact that it causes an effect does not depend on external circumstances. It is consistent in its causality. For example,

suppose a participant in our study viewed a series of images about a man eating fruit that implies that that fruit, Fruit A, gave him a rash. If the participant decided that Fruit A causes the rash, that participant would be assuming causal invariance if they applied their new causal knowledge to future situations. This participant, assuming causal invariance, would respond “No” to the question “Would you eat Fruit A?” On the other hand, a participant who did not assume causal invariance would not necessarily assume that Fruit A was always responsible for the rash. This participant might believe that its causality was specific to this situation, dependent on factors like the person who ate it.

In this way, the causal invariance assumption is crucial to our study. If participants do not assume causal invariance, testing the impact of parsimony differences or any other manipulations becomes useless (McGillivray & Cheng). This is because if the participant does not assume causal invariance, they will not apply the causality of the fruit anywhere beyond the limits of the specific episode presented to them to learn the causal link between fruit and rash. This renders them unable to answer questions requiring a generalization about the fact that if you assume Fruit A causes the rash once, you assume it will cause a rash the next time you eat it as well. Without this assumption, our asking you which generalized causal explanation is better does not make sense, because you do not think there *is* a generalized causal explanation.

The study last quarter was a pilot to fine tune the experimental materials, and hopefully eventually expand further into a study involving MRI scans during the experimental process to see the underlying brain regions involved in these processes. This quarter, we used the results from the pilot to develop a questionnaire to collect data in order to improve the future study—particularly by attempting to emphasize the idea of causal invariance and prevent subjects from missing that crucial assumption.

Methods

Participants

Our participants were 48 students at UCLA participating for course credit.

Design

Last quarter, we used a within-subjects design to measure the effect of the independent variable of parsimony level on two dependent variables: one was causal judgment and one was art preference—these were defined and measured as the participants' self-report responses when asked which explanation they thought was better, or which artwork they preferred. We were looking for a correlation between the two. We operationally defined parsimony for causal explanations as the level of conceptual simplicity in the explanation. For example, an explanation would be more parsimonious than another if it achieved the same purpose while requiring fewer things to be true, or eliminating unnecessary extra statements present in the other.

This quarter, we developed an online survey in preparation for the next stage of the experiment.

Materials

The pilot study experiment required a computer program that displayed to the participants three different conditions: the causal condition, the causal control condition, and the art condition. These three conditions were presented in random order to each participant, but the control condition never came before the causal condition.

The art condition was a series of 18 sets of paired artwork. Each set consisted of a slide with one artwork, a second slide with a second artwork, and a third slide with them both side-by-side, and an instruction to the participant to compare them and select which one they found more

“visually appealing.” The artworks were black-and-white computer-generated fractal images. One third of them had dimensionality level 1.2, one third had 1.5, and the rest had 1.8. Each image pair was one of three types: low-high (one 1.2 image, one 1.8 image), medium-high (one 1.5 image, one 1.8 image), or low-medium (one 1.2 image and one 1.5 image). The purpose of this section was to determine what level of dimensionality (or parsimony) the participant appeared to favor.

The causal reasoning section consisted of a series sets of slides each depicting a short chain of events. The first slide of each set was labelled “Day 1” and contained a picture of a fruit and a picture of a man, who either had or did not have a rash on his face. The next slide was “Day 2” which had either a picture of a different fruit, both the first fruit and a different fruit, or no fruit, and the man who either again did or did not have a rash. For example, the episodes might have occurred thusly: on Day 1, he ate Fruit A and had no rash. On Day 2, he ate both Fruits A and B, and did have the rash. The next slide reproduced the two episodes and then offered one explanation for the events (i.e. Fruit A does not cause the rash, but Fruit B does). The following slide had a different explanation (i.e. neither fruit causes the rash alone, but something special about the combination of the two of them causes the rash). For each account, the slide asked the participant to answer if that account explained the circumstances (yes or no). The final slide of the set asked the participant to choose which of the two was the “better explanation.” The purpose of this section was to assess the participant’s use of parsimony in causal reasoning—whether they would choose the more parsimonious explanation (in the example given, it would be the former, not the latter.) The specific manipulation in this section of the experiment was the level of parsimony; all other aspects of the explanations (like syntactic simplicity) were held constant. Each explanation pair was ranked on a scale of low, medium, or

high parsimony, and paired in the same ways as the art section: there was an equal number of low-high, low-medium, and medium-high pairings. There was also an additional condition—a control, where an explanation that did explain the circumstances was compared against one that did not (in the previous example, this might say “Fruit A causes the rash and Fruit B does not.”) This was to ensure that participants were actually taking into account the validity of the causal explanations.

The causal control section was the same as the causal reasoning section, save for two main differences. The first was that the question asked was not “which account is the better explanation,” but “which account do you prefer?” This was to eliminate any explicit causal thought process in the participant so that they could focus completely on the structuring of the accounts put in front of them. The second difference was that this time, instead of holding simplicity constant and varying the parsimony levels, there were five separate conditions with different variations between accounts. The five conditions of this were: simplicity change, parsimony change, congruent, incongruent, and a control. Simplicity change meant the parsimony level of the two conditions were the same, but the syntactical simplicity was different. We defined syntactic simplicity as the use of active voice as opposed to passive voice to say the same thing. For example, saying “Fruit A causes the rash” would be syntactically simpler than saying “The rash was caused by Fruit A.” Parsimony change meant simplicity remained constant, but parsimony level changed. For example, both sentences might be in active voice, but one would be the low parsimony level, while the other would be high. The congruent condition was when the parsimony level and simplicity level both changed, but they agreed for each explanation. For example, one account would be high parsimony in active voice, while the other would be low parsimony (high complexity) in passive voice. The incongruent condition was the

opposite of this: parsimony and simplicity disagreed with each other. That is, one explanation could be active voice with low parsimony, while the other would be passive voice with high parsimony. Finally, the control was like the control for the causal section: both accounts were in active voice, but one explained the situation and one did not. The purpose of all these conditions was to detect patterns in the types of answers that participants were giving, and to see what they were taking into account besides parsimony level to make their decisions.

The survey we developed this quarter resembles the causal control and causal reasoning sections of the pilot study, but with some significant format, content, and wording alterations. The purpose of this survey was to make sure that our pilot study was measuring what we wanted it to be measuring—and if it was not, to fix that.

Because the new survey is not meant to be a full-blown study, it is not incorporated into the full three-segment program of the pilot study. Instead, it is simply an online survey. The survey consists of seven blocks—one for instructions, and then one experimental block for each of the following conditions: blocking, confounding, double blocking, learned irrelevance, prevention, and catch. Each experimental block includes a concise version of the slides in the pilot study: both daily episodes presented together, followed by a new question that was not in the pilot, followed by both explanations presented side-by-side above a 6-point Likert scale asking the user to decide to what extent they feel one account is better than the other.

The blocking condition consists of a scenario of two fruits; call them Fruit A and Fruit B. The first episode shows that a character Bob ate Fruit A and got a rash. The second episode shows that Bob ate both fruits A and B, and got a rash again. The idea of “blocking” here refers to the fact that once we see that Bob had a rash after eating the first fruit alone, we assume that fruit A is what caused the rash. So when we see both A and B eaten together, we already have

the assumption that the rash-causer is A, even though B could additionally cause the rash as well. In this way, A “blocks” B. The two explanations provided for these episodes differ in parsimony—for example, one explanation will say that it is only A that causes the rash, while the other says both A and B cause it.

The confounded condition is just as it sounds—the solution to which fruit caused the rash is confounded. For example, in the first episode, there could be no fruits eaten, and no rash. Then, in the second episode, there is both Fruit A and Fruit B, and the rash is present. The participant has no way of knowing which fruit actually caused the rash: they are confounded. Thus, there is no difference in accuracy between the two possible explanations; just parsimony.

The double blocking condition is the same as the blocking condition, except for the presence of an extra fruit in the second episode. Thus, the first episode is identical to the blocking condition, but the second episode has three fruits instead of two. The purpose of this block is just to see if the extra fruit makes a difference in the way participants answer the question. It is mainly just for our own knowledge, and not intended to be one of the conditions in the real study.

The learned irrelevance condition’s first episode shows us Fruit A only, and a man called Larry who does not have the rash. We can conclude that A does not cause the rash. It is “irrelevant.” Then in the second episode, we see Fruit A and Fruit B, and Larry does have the rash. The participant must again decide between two explanations that both explain both episodes but differ in parsimony; the defining feature of this block is that, when thinking most parsimoniously, Fruit A does not need to be in the explanation at all.

The prevention condition also features two explanations that differ in parsimony level. The only difference is that while all the other conditions use causal language—they say that Fruit

A “causes”—the prevention condition uses preventative language—they say that Fruit B “prevents” the rash. The purpose of this section is to see if participants think differently in the context of prevention than causation.

The catch condition is the final condition, and its purpose is simply to make sure the subject is paying attention and not just providing random answers. It accomplishes this by supplying one valid explanation and one explanation that is clearly incorrect; it is inaccurate in explaining the data. If a participant chooses the wrong answer in the catch condition, it alerts us to the possibility that that participant was not actually doing the study as they were supposed to.

Using these conditions, we developed 15 different versions of the survey in order to completely counterbalance all the variables. All combinations of parsimony levels (1-2, 2-1, 2-3, 3-2, 1-3, 3-1) are present an equal amount of times across the versions and across conditions.

We also made some important content and wording changes to the questions and statements presented to the participants. Our feedback from the pilot study indicated some confusion about the meaning of our questions and explanations—are we making judgments assuming we know about the episodes, or are we pretending we are Bob before he eats any fruits? Is the question specifically for Bob, or do our personal thoughts come into the picture? And how do we really know what causes the rash? To try to eliminate the need for these questions, we added a new question between the presentation of the episodes and the presentation of the explanations. One example of the question is this: “Consider what you would do if you were Bob. Suppose you really liked the taste of [Fruit A]. Would you eat [Fruit A]?” We hoped that this question would get the participant to think about the problem in a specific way—to wonder what the causal relationship between the fruit and the rash is. In understanding an answering this question, they would be forced to think about whether they believed the fruit

caused the rash or not, and then weigh the options (enjoying the taste vs. getting a rash). Thus, it is not as important to us what their actual answer to this question is (whether they decide to eat it or not) because that requires personal input that is irrelevant. The importance of the question is that they engage in the thought process to evaluate the potential causal relation. Once the participant goes through that thought process, it will be much easier and clearer for them to answer the next question—the rating of which explanation is better.

The second major change we made is to remove the word “episode” from the explanation comparison section. This is because our feedback from the pilot indicated that the word was confusing participants. Since “episode” has a diagnostic, single-event kind of connotation, they were sometimes unclear as to whether our questions meant for them to look for a causal diagnosis specifically for the current situation only, with the current person, or whether they were supposed to think about the fruits’ causal properties in a general sense. We wanted to remove the possibility that the participant would fixate on the word “episode” and then make decisions in the context of only Bob, by assuming that the rash was specific to Bob, or that Bob’s opinions were relevant. By completely removing the word “episode” and rewording the questions so that no such noun was necessary, we hope to reduce participant confusion and distraction from the point of the slides.

Both of these changes—the extra question and the elimination of “episode” were in an effort to emphasize the causal invariance assumption. By focusing on the thought process of applying their causal deductions to the future, and by being rid of words that make them feel limited to the specific situation, hopefully the future participants will make the causal invariance assumption and go through the survey accordingly. If it turns out that these modifications do

indeed work, then we can use the data gathered to inform the wording and structure of the final study.

Procedure

In the pilot study, participants came into the lab one at a time. We handed them the consent form and asked them to read it. Then we led them to the computer and instructed them that all the instructions were on the computer. They then sat through the experiment, pressing buttons labelled “L” and “R” (for “right” and “left”), or “Y” and “N” (for “yes” and “no”) depending on what the question asked (when choosing between explanations, it was R or L, and when being asked if the account explained the episodes, it was Y and N). When the three conditions were over, the participants were asked to fill out a post-experiment questionnaire about their thought process during the experiment before they left. We used this questionnaire to help develop the survey this quarter.

We have not implemented the survey yet.

Results

From analysis of the pilot’s results, we discovered a few things. First of all, from the questionnaire it was apparent that there was some division between participants; while some felt that a more general explanation and a lower dimensionality picture was better, others felt that a more specific explanation and a higher dimensionality picture was better. On a positive note, however, these preferences did seem to correlate—if the participant said they made decisions based on greater causal detail, they also said they chose paintings with greater detail (high dimensionality). However, this difference in preference may lead to some ambiguous results for the tests in comparison of levels of parsimony, if participants had different ideas of what parsimony is.

Because of this ambiguity surrounding the idea of parsimony, the survey we developed this quarter heavily involves parsimony. All block types involve parsimony change, so that we can obtain results for a rich and detailed analysis of different ways people can think about parsimony.

Additionally, results from the pilot study as well as the end-of-study questionnaire indicated that there was some need to clarify what exactly we meant by our questions. In a few conditions, the results did not match up to the expected patterns, and participants reported confusion or varied interpretations of our questions. We also strove to resolve this, or at least improve it, with the work this quarter.

References

- Hawking, S., & Mlodinow, L. (2010). *The grand design*. New York: Bantam Books.
- Liljeholm, M., & Cheng, P.W. (2007). When is a cause the “same”? Coherent generalization across contexts. *Psychological Science*, *18*, 1014-1021.
- McGillivray, S. & Cheng, P. Analytic versus Empirical Knowledge of the Concept of Causal Invariance.
- Taylor, R. (2002). Order in Pollock’s Chaos, *Scientific American*, 116-121.

Virtual Reality and Cognitive Mapping Strategies In Language Acquisition

Taylor Henry

Introduction

For many students, learning is more about memorizing bits of information for the next exam - only to forget it immediately afterwards – and less about actually understanding and making connections to the material. Especially when becoming familiar with a new language, students need the skills to delve beyond surface level knowledge, understanding both how specific words fit into the broader language framework and what they mean. In order to gain mastery or become fluent, they need to be able to know the language in its own right, not just as a constant translation of their native tongue. Virtual reality provides a way to make this possible, or at least easier, than traditional lecture and memorization centered learning. Using task-based activities and games to connect what would, in a typical setting, just be a word and its definition on paper, to a virtual representation of the object, as well as its context within the language, creates a much more dynamic and natural learning environment (Chen & Su, 2011). Representing these words as objects in a virtual environment and placing them in a way that is consistent with their relationship to each other inherently utilizes cognitive mapping and the spatial learning method by taking advantage of something the brain does automatically (Johns, C., 2003), mentally mapping and storing information about a new place, and aligning this with what could be described as a vocabulary flow chart.

Method

Design

The experiment will attempt to teach participants 25 words of a language and test two independent variables, a learning method and a teaching method. The participants will be randomly assigned to one of three groups: one that is taught in a standard from-the-textbook style, one that participates in the virtual environment and is guided during the learning sessions, and another that participates in the virtual environment but is not directed during the learning sessions. The goal will be to first determine whether the use of cognitive mapping in VR facilitates learning and retention as well as the standard textbook method, and if so, whether being guided in study is more effective than allowing the participant to be responsible for their own learning.

Materials

- a. Lectured Learning: A classroom, along with a projector and blank flash cards.
- b. Virtual Environment: A lab room with two computers capable of running OpenSim virtual environments, as well as a joystick and two headsets will be required. The necessary virtual environments will include two different, but very basic houses. Both should have four rooms (a living room, kitchen, bathroom, and bedroom), each containing five household objects placed in various locations, preferably requiring user interaction to access (i.e. a saucepan inside a kitchen shelf that the participant must click on to open). The purpose will not be for the objects to be difficult to find, but the number of items in each room should be enough so that the participant will likely have to know what they are looking for in order to select the correct object. Upon clicking on an object, the spelling of its name should appear on the screen, and a recording

should read it out loud. Each room should contain a sign with the word for it written in the new language that can also be clicked upon to hear the pronunciation. The two houses should only be similar in that they contain the same rooms and the same set of objects. The layout of each should be different (longer hallways, different room sizes and arrangements, etc.), as should the filler objects in each room. The objects that are to represent vocabulary words should be slightly different between houses, but should be recognizable as the same type of object.

Procedure

Traditional Learning

Day 1: The students learning by traditional methods will participate in 45 minutes of guided study. A “teacher” will present a PowerPoint, each slide of which will feature a different vocabulary word, along with a picture of a corresponding object and its definition. This presentation will be followed by 30 minutes of individual practice, in which participants will be asked to make and study flash cards for each of the 24 objects, simulating one of the most popular methods for learning basic vocabulary words. At the end of this time, the participants will be asked to complete a quiz in which the teacher reads each new word in random order, and the participants are expected to write down the English translation. After the quiz, the flash cards will be collected. For all traditional quizzes, one point will be awarded for each correct answer.

Day 2: A shorter version of the first day’s learning session, the teacher will quickly review the PowerPoint presentation, and then allow ten minutes for flash card study, after which the participants will be tested in the same manner as the previous day.

Day 9: The participants will return for one last quiz, this time without any review or additional time to study.

Guided Exploratory Learning

Day 1: The participant will be given a joystick orientation and fifteen minutes to create their avatar. Upon completion of this, they will be asked to teleport to the first house, which they will enter with the experimenter. During this first learning session, all of the test objects will be highlighted in some way, so they are easy to find in the time given. The experimenter will give the participant five minutes to explore each room, suggesting that they spend their time studying the test objects and listening to their pronunciations. After all the rooms have been explored, the first testing session will begin. The participant will be asked to find the object representing a given vocabulary word, and will be given the room location of the object, as well as the name of an object nearby. If the language being learned was Spanish, for example, each testing task would read like this: “Find la almohada (pillow) in el cuarto (the bedroom). It’s near el tocador (the dresser).” Each task will send the participant to a different room, and all twenty of the objects will be tested. Scoring will be based off of whether or not the participant found the object, how long it took them (1 minute time limit), and how many unnecessary rooms they visited before they entered the correct one. Following the same procedure as the first learning session, the second session will last a total of 15 minutes, and the participant will be allotted 3 minutes and 45 seconds in each room. The only change in the second testing session will be that the task will no longer give a second hint and will only reveal the name of the object and the room that it is in. The third learning session will last ten minutes (2 minutes and 30 seconds in each room), and the final testing session of the day will only give the name of the object to be found.

Day 2: The second day of the experiment will progress similarly to the first, except that it will begin with a fourth testing session (where only the name of the object is given), followed by a

ten minute learning/review session, and end with a fifth testing session no different than the fourth.

Day 9: The participant will be tested verbally, in an office. They will be given the English translation for each of the 24 words, and will be expected to answer with the learned language equivalent.

Free Exploratory Learning

The only difference between the procedures for the free and guided exploratory learning trials is that during free exploratory learning sessions, the participant is not instructed as to which rooms to visit. Hopefully, they will use this freedom to design their own “study plan”, and especially in later learning sessions, take advantage of the option to focus on the rooms and/or words they need more practice with.

Expected Result

While a previous study on guided versus free navigation found that those who were guided exhibited a better memory (spatial score) of the rooms and the objects in them than those who had free reign, this may have been due to the very limited amount of time participants could spend in each room (90 seconds). In that study, participants had no time to waste in examining each room, and those who were guided most likely benefited because they didn’t spend any time making choices as to which rooms to visit and for how long (Betella, A., Bueno, E.M., Bernardet, U., & Verschure, P., (2013). In the proposed study, however, it is predicted that those with free exploratory learning sessions will perform better than those with guided learning. Though it will take time for the free exploratory participants to make choices about what objects to study, the experiment design allows them to do so while also maximizing the potential benefits of creating an individualized plan.

Compared to the traditional teaching method, it is predicted that VR teaching will be at least as effective, and possibly more so when it comes to long-term recall. Even if the difference in results between VR and the lecture-style is negligible, VR can almost certainly be made more accessible, customizable, and engaging than standard educational strategies.

References

- Betella, A., Bueno, E.M., Bernardet, U., & Verschure, P. (2013). The effect of guided and free navigation on spatial memory in mixed reality. *Laval Virtual*.
- Chen, H.J. & Su, C.C (2011). Constructing a 3D Virtual World for Foreign Language Learning Based on Open Source Freeware. *Edutainment Technologies*.
- Johns, C., (2003). Spatial Learning: Cognitive Mapping in Abstract Virtual Environments. *Afrigraph '03*.

IMPORTANT NOTICE: The following is an undergraduate research paper submitted as part of the requirements for one quarter of participation in a research apprenticeship in the UCLA Human Perception Laboratory (HPL) in the Department of Psychology at UCLA. This document has not undergone a scientific review and may or may not relate to or correctly represent projects currently or previously undertaken in the HPL. Some features of adaptive and perceptual learning techniques mentioned or described herein are protected by US Patent #7052277 and patents pending. All rights reserved. For information, contact info@insightlearningtech.com.

Differences in Learning Across Criterion Trials in Adaptive Sequencing and Fixed

Expanding Schedules

Victoria Groysberg

The principles that govern successful learning when people learn new perceptual classifications and the ways in which the order of presentation of material can be optimized remain questions that require further scrutiny. Previous studies have shown that appropriately designed Perceptual learning (PL) technology can produce rapid and enduring advances in learning (Kellman et. al 2009). PL is the process of learning improved skills of perception. PL differs from other types of learning in that it can occur over a wide scale of time, like weeks, and it can occur for a wide variety of perceptual tasks, ranging from sensory discriminations to very tactile acuity tasks. PL is also restricted to the specifications of the stimuli and task where training is occurred, and the learners do not usually require feedback in order to exhibit PL effects. The PL often does not transfer to other tasks, stimuli, or sensory locations.

Adaptive schedules adapt the presentation of educational material according to students' learning needs, as indicated by their responses to questions and tasks. Adaptive schedules seem to outperform fixed schedules, even when fixed schedules are expanding, as measured by efficiency. Results of experiments in a dissertation study showed that an adaptive scheduling algorithm produced greater learning gains than fixed schedules. This was true for when the total number of presentations was limited the gains were measured after a one-

week delay. Adaptive schedules still outperformed fixed condition schedules, expanding schedules, in terms of learning gains at immediate and delayed tests (Mettler 2014).

In Adaptive schedules, learning strength could be reflected in response times. A previous study used the ARTS (adaptive response-time-based spacing) Algorithm for predicting item strength and appropriate spacing based on RTs. ARTS used both accuracy and response time (RT) as direct inputs into sequencing (Mettler & Kellman 2014). Response times were used to assess learning strength and to determine mastery, making both fluency and accuracy goals for the participants. The study used adaptive learning and created the spacing between items presented by expanding item recurrence intervals as an inverse function of RT. Third graders in an online school learned basic multiplication facts in about two hours using ARTS and outperformed a control group using standard instruction. Results were found to be positive and the algorithm resulted in increases in learning.

The usefulness of research into learning criteria in relation to spacing cannot be overstated. Adaptive schedules apply across many domains. In difficult domains such as Chemistry, adaptive schedules are perhaps more effective when learners learn to criterion. Whether expanding intervals of spacing align with the increasing learning strength of items is a vital question that was investigated in this study using Chemistry Nomenclature Experiments. What types of spacing make learning ideal, whether adaptive schedules result in expanding spacing, and when is expanding more beneficial was investigated. The hypothesis investigated was whether adaptive schedules resulted in greater increases in learning than continuous, fixed expanding schedules.

Method

Participants

The participants of this experiment consisted of 31 community college chemistry students, Collin College, Tx. No ethnic or demographic criteria were used to recruit participants. All participants merely met the requirements of speaking and reading fluent English.

Design

The experiment had a between subjects design and each participant was assigned whether they would take the Adaptive or Fixed Scheduling Module. The independent variable in this experiment was the type of scheduling used. Two levels for type of scheduling were used, fixed continuous expanding and adaptive. In both conditions, a Pretest was conducted, then the Training was conducted, followed by a Posttest in one session. Then a Delayed posttest was conducted in a second session 2 weeks after the participant started the module. Each learning item was presented 4 total times in both conditions. Items were presented to a learning criteria.

Materials and Apparatus

A web-based Perceptual learning module (PLM) was used, consisting of 36 chemistry nomenclature facts. The module tested Chemistry fact learning such as Ion naming and acid naming of both converting from a Formula to a Name and from a Name to a Formula. In the Continuous expanding fixed condition, the spacing between presentations was 1-5-9-12-15-18-21 etc. In the fixed condition, the learning session stopped when learners met a learning criterion for all items or all items presented 15 times. In the adaptive scheduling condition, the spacing of the items was based on the RT of the participant. In the adaptive condition, the learning session

stopped when the participant reached the learning criterion of responding with a 100% accuracy, answering 4/4 items correctly in a RT less than 20 seconds.

Results

Figure 1 presents the efficiencies of the continuous fixed condition schedule and of the adaptive schedule. As can be seen in Figure 1, it appears that in the Chemistry Nomenclature Study, adaptive schedules did outperform fixed schedules, even when fixed schedules were expanding, as measured by efficiency. The efficiency measure was defined as a change in accuracy (accuracy of responses in the post-test minus accuracy of responses in the pre-test) divided by number of trials conducted in the training phase. Figure 2 demonstrates RTs of accurate responses. The figure demonstrated that there were decreasing average RTs for correct trials in the adaptive condition but that RTs remained stable in the fixed condition. Figure 3 demonstrates that there was a higher proportion of participants who had perfect accuracy across the last 4 presentations of every item in the adaptive condition than in the fixed condition. 87% of the participants in the adaptive condition versus 18% of participants in the fixed condition of participants had perfect accuracy in the last 4 presentations. Figure 4 demonstrates that there was a lower average accuracy in the fixed condition than in the adaptive condition. Average accuracy increased but then plateaued in the adaptive condition. Figure 5 demonstrates that the average RT regardless of correctness was lower in the fixed condition than in the adaptive condition. Figure 6 demonstrated that the average delay between presentations was higher in the fixed condition than in the adaptive condition.

Discussion

Results of Figure 3 demonstrated what was valuable to see, that it is possible to have large spacing delays with accurate responses. This demonstrated that it was indeed possible to have fluent retrievals across widely spaced intervals. Figure 1 demonstrated that there were decreasing average RTs for correct trials in the adaptive condition but that RTs remained stable in the fixed condition. These results imply more change in learning strength in the adaptive condition than in the fixed condition, thus supporting the hypothesis that adaptive schedules are superior to fixed schedules. RTs remaining more stable in the fixed condition implies that in the fixed expanding condition, the learning expanded past the point of substantial recall.

Results of this study demonstrated that RTs remained more stable in the fixed condition than in the adaptive condition and that the learning expanded past the point of substantial recall in the fixed condition. A potential future study would be a variation of the Chemistry nomenclature studies by adding a constraint to the fixed condition so that the learning would never expand past the point of substantial recall in the fixed condition. Perhaps adding the criterion of an RT would correct this problem in a future study, a point of investigation.

Figure 5 demonstrated the unexpected result that average delay between presentations was higher in the fixed condition than in the adaptive condition. Figure 4 also demonstrated the unexpected result that the average RT regardless of correctness was lower in the fixed condition than in the adaptive condition but many other results did support the hypothesis that adaptive schedules promoted learning more than fixed schedules. Figure 1 showed that there were decreasing average RTs for correct trials in the adaptive condition but that RTs remained stable in the fixed condition. Figure 2 demonstrated that there was a higher proportion of participants

who had perfect accuracy across the last 4 presentations of every item in the adaptive condition than in the fixed condition.

Expected results may have occurred because participants in the fixed condition might have demonstrated lower learning because of decreased processing and less attention to situational variables due to greater fatigue effects. A fixed schedule may have fatigued participants more than an adaptive schedule, tailored to the needs of the participant did. Unexpected results may have occurred because in the adaptive condition, a higher cognitive load was created and more time to process the information was needed by the participant, thus resulting in a desirable difficulty that led to better learning and thus a higher accuracy.

The current study provided conflicting results so future work needs to be conducted because the role of the learning criteria in relation to spacing is a highly underresearched area but is a promising avenue in the search for a greater understanding of the interaction with spacing. The potential of changing training programs and improving educational systems using findings that can improve learning fuels further research into the use of adaptive schedule in different types of learning. The potential of allowing a greater access to an education by providing similar tools of learning to highly varying socioeconomic classes through web-based PLMs also fuels future research into how to improve learning modules.

References

- Kellman, P. J., Massey, C. M., & Son, J. Y. (2009). Perceptual learning modules in mathematics: Enhancing students' pattern recognition, structure extraction, and fluency. *Topics in Cognitive Science*, 2(2), 285-305.
doi:10.1111/j.1756-8765.2009.01053.x
- Mettler, E & Kellman, P. J. (2014). Adaptive response-time-based category sequencing in perceptual learning. *Vision Research*, S0042-6989(13)00301-5.
doi: 10.1016/j.visres.2013.12.009.
- Mettler, Everett William. University of California, Los Angeles, ProQuest, UMI Dissertations Publishing, 2014. 3605425.

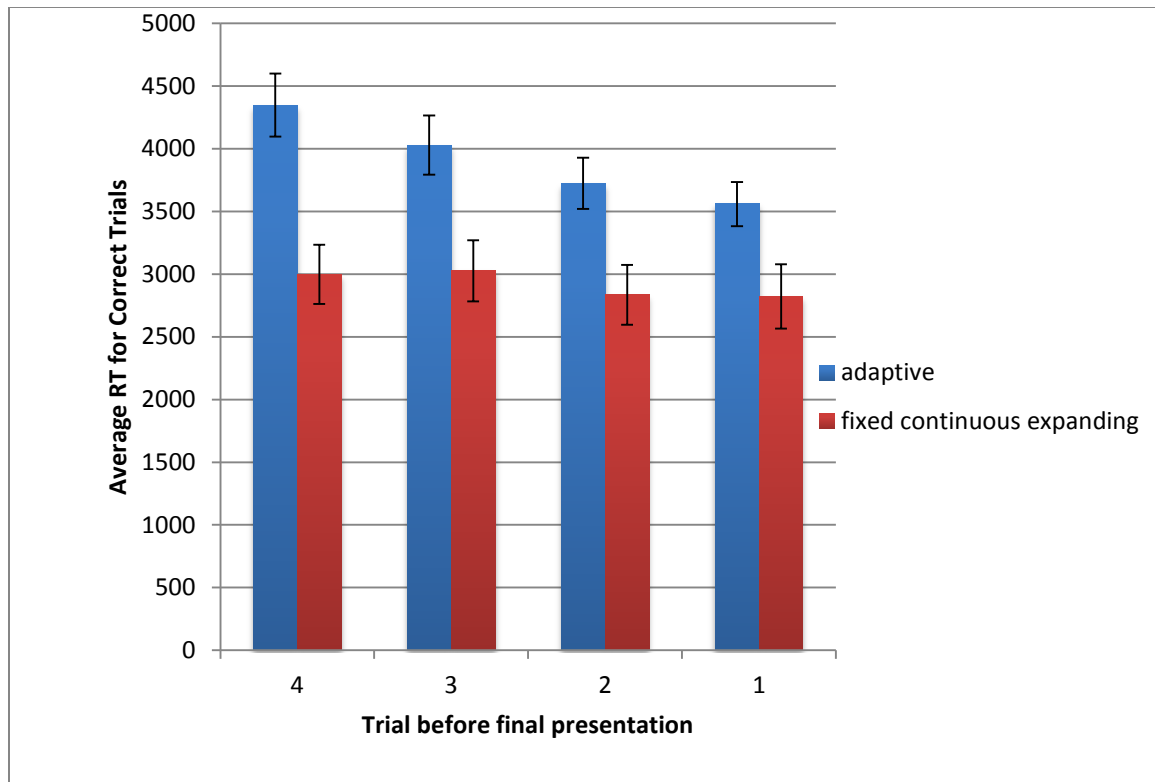


Figure 1: there were decreasing average RTs for correct trials in the adaptive condition but that RTs remained stable in the fixed condition.

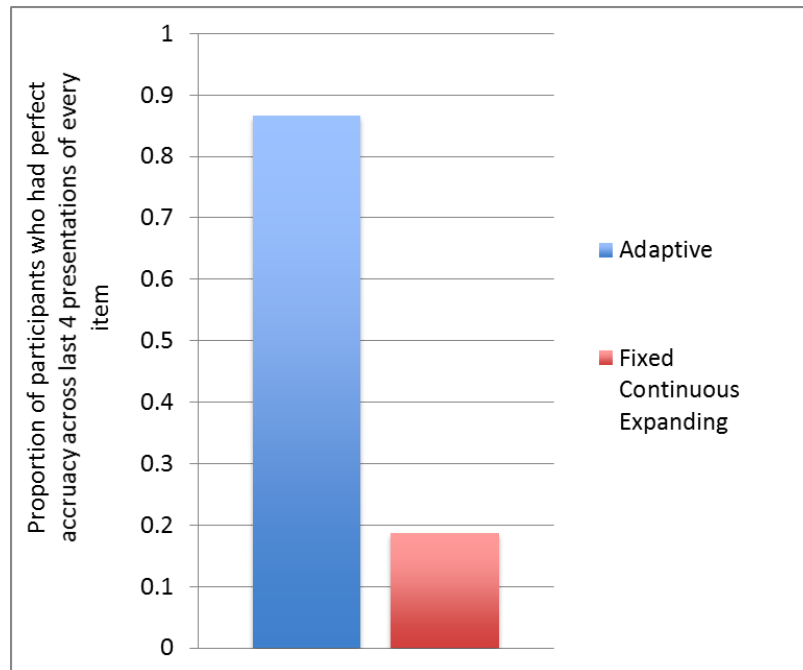


Figure 2: there was a higher proportion of participants who had perfect accuracy across the last 4 presentations of every item in the adaptive condition than in the fixed condition. 87% of the participants in the adaptive condition versus 18% of participants in the fixed condition of participants had perfect accuracy in the last 4 presentations.

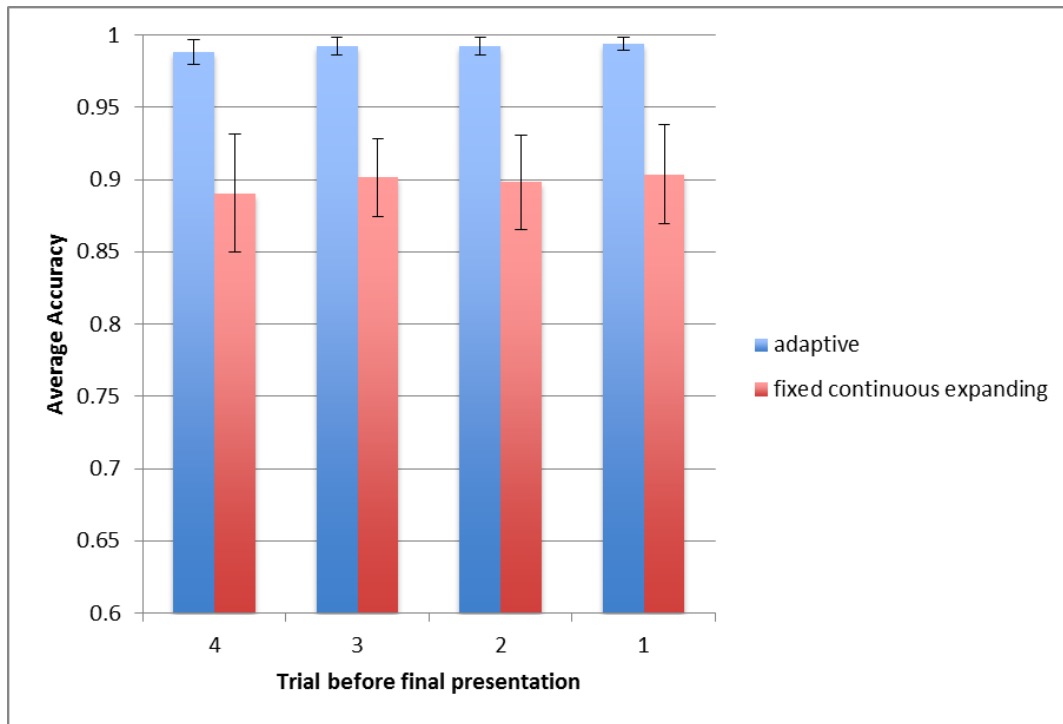


Figure 3: there was a lower average accuracy in the fixed condition than in the adaptive condition.

Average accuracy increased but then plateaued in the adaptive condition.

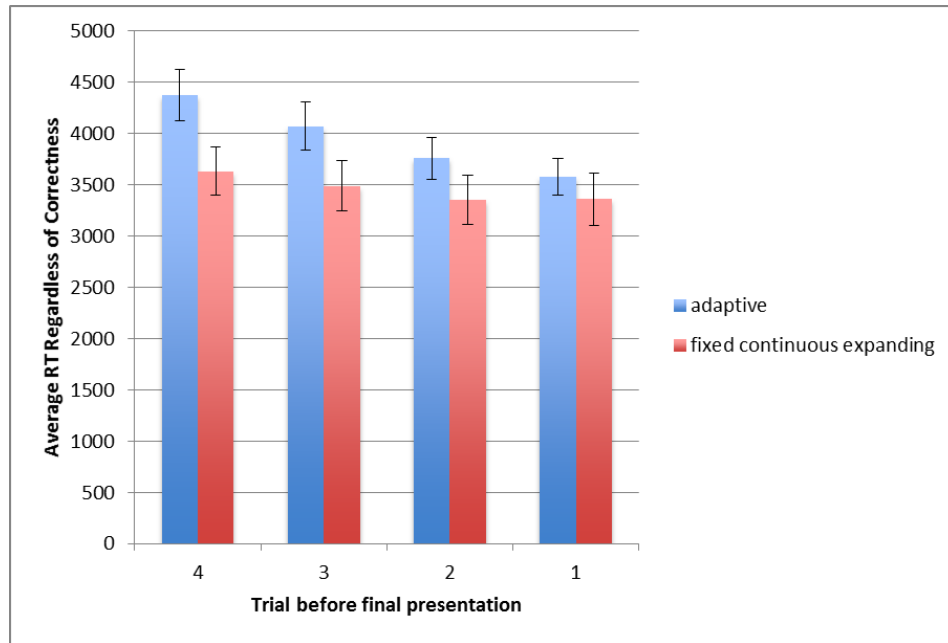


Figure 4: the average RT regardless of correctness was lower in the fixed condition than in the adaptive condition.

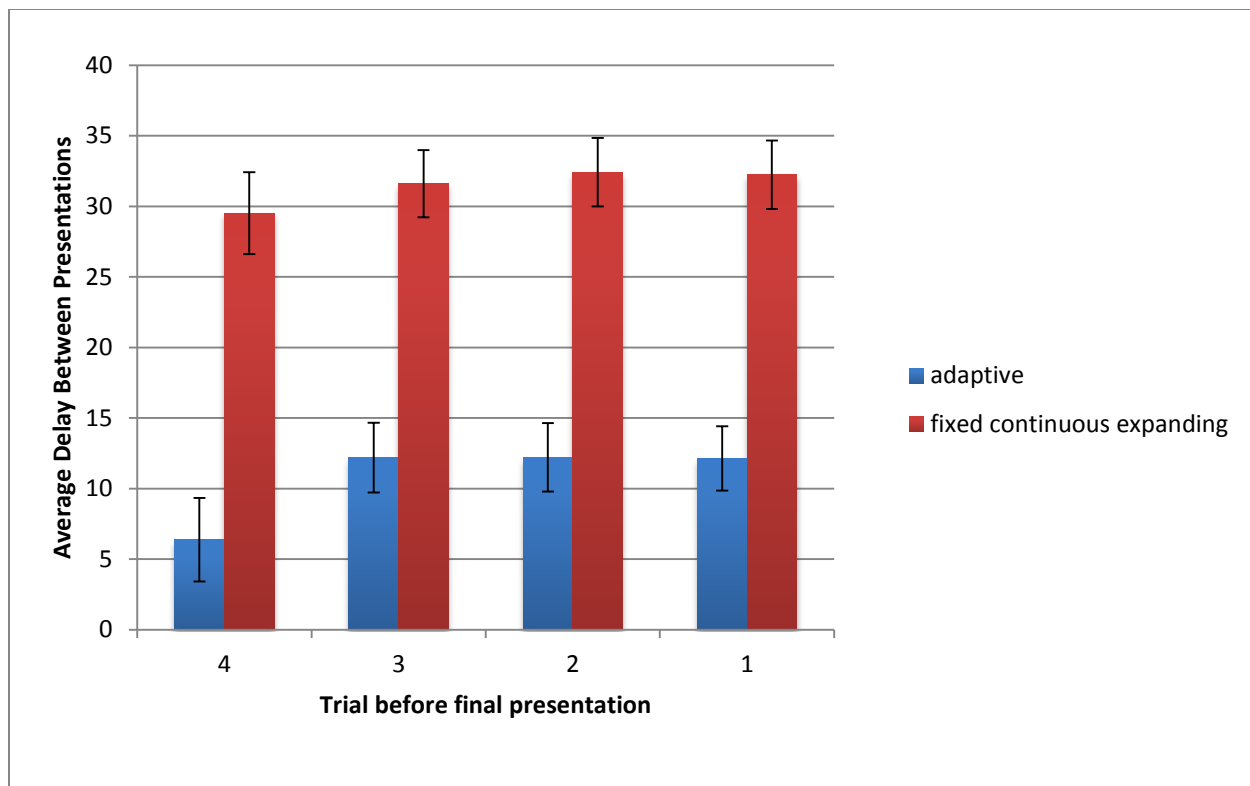


Figure 5: average delay between presentations was higher in the fixed condition than in the adaptive condition.