Gaze tracking variables as indicators of learning

Iyer, Arjun

http://hdl.handle.net/2144/15079

Boston University
GAZE TRACKING
VARIABLES AS INDICATORS OF LEARNING

by

ARJUN IYER
B.A., Bates College, 2011

Submitted in partial fulfillment of the
requirements for the degree of
Master of Arts
2014
Approved by

First Reader

Ann Zumwalt Ph.D.
Assistant Professor of Anatomy and Neurobiology

Second Reader

Mark Moss, Ph.D.
Professor and Chair, Anatomy and Neurobiology
DEDICATION

I would like to dedicate this work to my family and my friends for their support, as well as the faculty at Boston University for their continued involvement in my academic career.
ACKNOWLEDGMENTS

I would like to acknowledge my principle investigator and first reader, Dr. Ann Zumwalt, for all of her involvement. I would also like to thank my lab partners for their help in data analysis and conducting the experiment.
GAZE TRACKING
VARIABLES AS INDICATORS OF LEARNING

ARJUN IYER

ABSTRACT

The process of learning contains multiple aspects, whose intricacies have yet to be fully understood. The current experiment utilized gaze tracking technology to observe whether variables in subjects’ gaze (e.g., total time spent on a given image, percentage of time spent on cognitively salient features of the image) was predictive of the subject learning the material. Subjects consisted of students from a Medical Gross Anatomy course, who were tested twice—once before they had learned the course and once after they had learned a certain amount of material. Following the baseline testing, subjects were broken up into three groups, A, B and C, each representing a later visit in the course. Results indicated that groups B and C tended to spend more time on cognitively salient areas of interest, but this was moderated by familiarity. Moreover, results indicated that groups B and C tended to spend more time on images overall (end time) compared with group A or the baseline group. Overall, the results obtained were ambiguous, and warrant further study in order to arrive at a clear conclusion. Future directions of study may want to consider other gaze tracking variables, such as the time at which a subject first looks at a cognitively salient area of interest.
# TABLE OF CONTENTS

A detailed table of contents is provided for the document, listing all sections and their page numbers for easy reference. This helps in navigating through the content of the document efficiently.
Dwell Time/tracking ratio: .................................................................15

Cognitively Salient Area of Interest fixation time (CogaoiFT): .....................16

End Time: ............................................................................................16

Obscure images: ....................................................................................17

Analysis: ................................................................................................18

CogaoiFt analysis.....................................................................................18

End Time Analysis...................................................................................18

Correlations............................................................................................19

RESULTS .................................................................................................20

Cognitively salient AOI analysis:.............................................................20

End Time Analysis: .................................................................................23

DISCUSSION ..........................................................................................34

APPENDIX...............................................................................................40

Examples of images from the Back and Limbs section, Thorax Abdomen Pelvis

Section, and Head and Neck sections respectively.....................................39

REFERENCES.........................................................................................43

CURRICULUM VITAE..............................................................................45
LIST OF FIGURES

Figure 1a: Mean percentage of fixation time on cognitively salient areas of interest (CogaoiFT-measured as a percentage) as a function of time spent in the course (GROUP). ..............................................................24

Figure 1b: Mean percentage of fixation time on cognitively salient areas of interest (CogaoiFT-measured as a percentage) as a function of time spent in the course (GROUP). .........................................................24

Figure 2: Mean percentage of fixation time on cognitively salient areas of interest (CogaoiFT-measured as a percentage) as a function of time spent in the course (GROUP) with respect to obscure images only. ....................................................25

Figure 3a: Mean percentage of fixation time on cognitively salient areas of interest (CogaoiFT-measured as a percentage) as a function of time spent in the course (GROUP) with respect to back & limbs images only. .............................................25

Figure 3b: Mean percentage of fixation time on cognitively salient areas of interest (CogaoiFT-measured as a percentage) as a function of time spent in the course (GROUP) with respect to thorax abdomen pelvis images only. ........................................26

Figure 1x: For participants with an MRT score of 40 or below: mean percentage of fixation time on cognitively salient areas of interest (CogaoiFT-measured as a percentage) as a function of time spent in the course (GROUP). .................................26

Figure 1y: For participants with an MRT score of 60 or above: mean percentage of fixation time on cognitively salient areas of interest (CogaoiFT-measured as a percentage) as a function of time spent in the course (GROUP). .................................27

Figure 2x: For participants with an MRT score below 40: Mean percentage of fixation time on cognitively salient areas of interest (CogaoiFT-measured as a percentage) as a function of time spent in the course (GROUP) with respect to obscure images only. ..........................................................27

Figure 2y: For participants with an MRT score above 60: Mean percentage of fixation in obscure images only. ........................................................................................................28

Figure 3ax –for participants with an MRT score below 40: Mean percentage of fixation time on cognitively salient areas of interest (CogaoiFT-measured as a percentage) as a function of time spent in the course (GROUP) with respect to back & limbs images only. ........................................................................28
Figure 3ay – for participants with an MRT score above 60. Mean percentage of fixation time on cognitively salient areas of interest (CogaoiFT-measured as a percentage) as a function of time spent in the course (GROUP) with respect to back & limbs images only. .................................29

Figure 3bx – for participants with an MRT score below 40: Mean percentage of fixation time on cognitively salient areas of interest (CogaoiFT-measured as a percentage) as a function of time spent in the course (GROUP) with respect to back & limbs images only. .................................29

Figure 3by – for participants with an MRT score above 60. Mean percentage of fixation time on cognitively salient areas of interest (CogaoiFT-measured as a percentage) as a function of time spent in the course (GROUP) with respect to back & limbs images only. .................................30

Figure 4a: Total time spend on each image (End Time-measured in milliseconds) as a function of group (A, B, or C). Participants in groups A, B, and C had recently completed the back and limbs, thorax abdomen pelvis, and head and neck exams respectively. .................................30

Figure 4b: Total time spend on each image (End Time-measured in milliseconds) as a function of section (BL, HN, O, or TAP). .............................31

Figure 5: Comparison of fixation on cognitively salient areas of interest with the baseline group included (base). .........................................................31

Figure 6: comparison of end time as a function of group, with the baseline included. .................................32

Figure A: correlation of students' MRT scores with their final lab grades.................................................................32

Figure B: correlation of students’ MRT scores with their final practical grades. .................................33
INTRODUCTION

Learning

Learning is a multifaceted process essential for growth and development whose intricacies have yet to be fully understood. From academic success and skill building to social awareness and societal functioning (Judd 2006), the importance and ubiquity of learning is evident. The learning process recruits several distinct neurological phenomena, which themselves merit research; examples include selective attention, rote memorization, and executive function. These are recruited to varying degrees when one learns, depending on the task involved. For example, learning seemingly random facts, such as word lists and number strings, recruit areas such as the hippocampal formation, which is implicated in rote memorization. On the other hand, skills requiring procedural memory, such as learning to ride a bike, involves the cerebellum and its cortical modulatory connections (Draganski 2006).

From a practical standpoint, a more thorough understanding of learning could lead to methods that accurately measure the progression and culmination of this process. Academic institutions typically measure learning through exams or other indirect forms of assessment, a process that typically culminates in the form of grades. Grades are not without merit; they are a useful method of standardization in the application process, and are a good predictor of future performance. Students who obtain good grades in high school are more likely to perform well in college than students who perform below average (Brookfield 1986, Michaels 1989).
However, grades are limited in what they can actually address in the learning process. One reason for this is that grades are the end result of a progression, rather than indicative of the process itself. For a given topic, two students may wind up with similar grades, but may have used different techniques to learn the material. One student may have committed the knowledge to short term memory for the sake of the assessment, while the other student may have attempted to cultivate a more thorough understanding of the topic, presumably allowing for the knowledge to be retained for a longer time than in the case of the former student (Meltzoff 2009). It is also possible that students may have had extensive experience in a given topic beforehand, and therefore actually learned very little throughout the course despite their high marks. Because grades do not account for individual differences in learning techniques, they may not be the best guide for a professor to modify his or her material in order to accommodate the students. Finally, although good grades in one level of education may predict success in higher educational institutions, this correlation does not necessarily speak to the skill that is being learned; rather, it is a reflection of the similarity in assessment methods across levels of education (Knowles 1970).

Thus, researchers currently face the challenge of finding a method that can accurately measure the process of learning in addition to the end result. In experiments used to monitor the neurological processes in learning, such as memorization and attention, researchers have typically used instruments that measure brain activity in different areas, such as fMRI or EEG (e.g. Mulerf 2004). These methods are useful in
assessing the neurological change that occurs in learning, but they do not always correlate with an observable behavioral change (Archambault 1999, also see: Schyns 2011).

**Eye Tracking**

Although not created specifically for studying learning, eye tracking software is versatile in its applications. Alfred Yarbus, a Russian psychologist, was among the first researchers who investigated the applications and implications of gaze tracking variables. In his seminal work “Eye Movements and Vision”, Yarbus monitored the way participants tracked a scene as a function of the instructions given. From the study, Yarbus determined several gaze tracking variables that were indicative of how subjects perceived a scene. One example of such a variable is fixation time and location; Yarbus found that subjects’ eyes fixated on different areas of a scene if they were asked a specific question about the scene, such as the location of an object. These fixations differed when Yarbus asked participants to passively view the scene (Yarbus 1967). Thus, from analyzing variables such as fixation duration and saccade length, Yarbus was able to determine the intent behind subjects’ viewing a scene, as well as the relative salience of certain features in an image.

Through his experiments, Yarbus determined several noteworthy variables in eye tracking, such as saccade length and frequency. A saccade is a quick movement toward a particular stimulus; thus, by studying saccades, Yarbus wished to know how a subject’s eyes moved across features of a given image as a function of the experiment parameters.
Another variable of interest for Yarbus was fixation time, which is defined as the length of time a subject spends on a particular feature of an image (Yarbus).

The work of Yarbus spawned a new field of study investigating the applications of eye tracking. Several subsequent studies indicated that gaze tracking patterns in a given image can also differentiate subjects’ relative expertise in a given subject. For example, Matsumoto and colleagues (2011) performed a study in which they had subjects examine CT images of the brain. The subjects consisted of 12 neurologists (experts) and 12 novice controls (individuals with minimal experience in neurology). Results indicated that novices tended to fixate on visually salient areas, regardless of clinical importance. However, neurologists, though attentive to visually salient areas in the brain, also fixated on more cognitively salient areas that served as clues for diagnosis. Thus, Matsumoto and colleagues distinguished between “bottom up” scanning, in which one has no knowledge of the scene and looks for visually salient areas, and ‘top down” scanning, in which one has prior awareness of the image and can modify his or her gaze patterns accordingly. Thus, it would appear that although visual salience affects both novices and experts, the latter tend to focus on areas that are not explicit in a given image but important for the task at hand (also see: Savelsberg et al., 2002).

The difference between novices and experts was also addressed in a paper by Beverly Wood (1999). Wood, like Matsumoto, proposes that visual experts have a pre-conceived awareness of the scene, and so their scanning patterns reflect a knowledge beyond the objective features of the image. Wood focused on the gaze patterns of radiologists, whose field requires them to become familiar with various anatomical scans. Thus,
Wood also alluded to the ‘top down’ vs ‘bottom up’ approach, proposing that expert radiologists utilized the former process to accurately come up with a diagnosis. Finally, Wood noted that experts, although they spend less overall time on the images than novices, tend to focus for a longer period of time on the cognitively salient features of the image (also see: Kocak 2005, Harvey 2014).

**Current Experiment**

Since noticeably different gaze patterns exist between novices and experts, it may be possible to assess the progression and culmination of learning through comparing the differences in gaze tracking patterns within a given group over time. Using gaze tracking as an indicator of learning would provide a more detailed insight into how the student is progressing throughout the class as well as the student’s thought process when approaching a problem—two pieces of information not represented by number or letter grades.

Gaze tracking may also differentiate between rote memorization and analytical reasoning when a student approaches a question. If a student is asked to identify and image, he or she may do one of two things to arrive at the answer: one possibility is that the student may have simply memorized the image from his or her studies, and therefore would not need to look at the image actively. The other possibility is the student is not familiar with the particular image, but is familiar enough with images of that nature to use problem solving skills and analytical reasoning to arrive at the answer. In this case, the latter student may spend more time looking at cognitively salient features of the
image that provide clues to the student. Both students may ultimately arrive at the same answer, but through gaze tracking one may gain more insight into the method used (Lai et.al, 2013, Nodine et al., 1999).

The current experiment will begin assess whether gaze tracking variables can accurately assess the progression and outcome of learning. We tested participants at different periods throughout a gross anatomy course—one at baseline, and once after the students had learned a particular anatomy topic. The experimenters asked students to identify particular structures on dissection images while monitoring the gaze patterns of students We hypothesize that students will spend more time on cognitively salient features of a given images after learning the corresponding topic compared to before learning the topic. We measured fixation time on cognitively salient images as a relative percentage of the total time participants spent on the image. This hypothesis is based upon the findings that experts in a given field, when attempting to identify an image, will exhibit gaze patterns that that are not based on salience so much as useful information about the image (Matsumoto, Wood). Each

The other hypothesis concerned end time, defined as total time spent on the image, regardless of the gaze patterns. Since the experiment was constructed into discreet sections, it was hypothesized that students would not spend as much time on images they did not know, since they would be aware that they had not yet learned the section. Thus, a main effect of end time was expected such that groups that we tested later would spend more time overall on the images.
The experiment also contained obscure images of animal dissections in order to assess general problem solving ability—the students would not have seen these images before, so rote memorization would not be possible. For these obscure images, it was hypothesized that with each successive group (i.e. more time spent in the course), there would be increased fixation on cognitively salient images. We also hypothesized that there would not be a significant difference in end time, since students would be equally unfamiliar with the images, regardless of when we tested them.
METHODS

The protocol of this experiment was approved by BUSM IRB protocol H-32308

Participants: Participants consisted of 31 graduate students from Boston University who were enrolled in a Medical Gross Anatomy course. Prior to the start of the course, the experimenters sent the incoming class an e-mail informing them of the study (class size: 175 students). The e-mail contained the approximate timeframe of the experiment, as well as exclusion criteria; both of these will be discussed below. Participation was voluntary, and subjects were compensated for their time. The experimenters assigned participants a random subject number to retain their anonymity. Inclusion criteria required that students have normal or corrected vision (subjects were asked to wear glasses during the gaze tracking if they did so normally), be right handed, and not previously taken a gross anatomy course.

Recruitment and timeframe: Participants were informed via e-mail that they could sign up at any of the times assigned for the experiment using a Google document. The experimenters sent participants an e-mail informing them that they would be visiting three times to complete the experiment and that they would be subjected to EEG administration and gaze tracking. The gaze tracking experiment was the second part of a series of three experiments, the first and third of which involved EEG. These will not be discussed since the analysis and results are based only on experiment 2 (eye tracking with no EEG). All subjects took the baseline version of the task before they had started their anatomy course. The baseline task took participants approximately an hour and a half,
since it contained all three experiments—the first and the third of which involved EEG, and the second of which only involved gaze tracking. This thesis will discuss the second experiment only. The participants were then divided randomly into three groups of ten—groups A, B, and C. These groups took experiment two (just gaze tracking; results below) once again following one of their three exams: group A completed the task again following the Back and Limbs exam in anatomy (exam 1), group B following the Thorax Abdomen Pelvis exam (exam 2), and group C following the Head & Neck exam (exam 3). Participants were sent a second e-mail informing them they could sign up for the second part of the EEG experiment once the class had ended, but before winter break. During the participants’ second visit, unlike the baseline group, the participants were not set up for the EEG experiments. In addition, during participants’ second visit the experimenters added seven animal dissections which the baseline group had not seen (see procedure for rationale behind the images).

Location: The experiment took place in the Spivak Center located on the Boston University Medical campus. The experimenters allotted one room for consenting and EEG set up, and used another room to administer the three tasks.

Materials

The experiment was conducted using SMI software, which included a gaze tracking camera (Red-m) and SMI (SensoMotoric Instruments) laptop. The SMI company specializes in eye-tracking software which may be used in a variety of applications (see website: http://www.smivision.com). The gaze data was collected from the camera via Iview, a specialized software, which then transmitted the data to BeGaze.
Experiment Center is a software program provided by SMI in which one may construct a gaze tracking study using image and text files. The experimenter chooses which pictures to add from files downloaded or created elsewhere; text files may be created directly through Experiment Center. The experiment was constructed as follows: the first slide consisted of instructions on the experiment (see below for details on procedure). Afterwards, an image slide was shown with an arrow followed by a text slide, asking participants to identify the structure indicated by the arrow and allowing participants to type their answers. Two experiments were created via experiment center: one experiment for the baseline visit, and one experiment for the participants’ revisits. The second experiment included obscure images. Once the experiment was created, the experimenters administered the task by first locking the software, allowing no further changes to be made. The experimenter then hit “record”, whereupon a screen appeared prompting the experimenter to enter the name of the data to be collected. On this screen, the experimenter entered the subject’s four digit Subject Number.

Once the participants completed the experiment, the experimenters analyzed the results obtained using SMI’s BeGaze software. Through BeGaze, one may open an experiment that was created in Experiment Center and analyze the gaze tracking data of each subject for each stimulus. The BeGaze software displays various gaze tracking variables, which will be discussed below in detail. BeGaze also displays the subject’s scan path—the order in which the subject fixates on areas of the image and for how long—and a gaze replay—a real-time representation of how the subject’s gaze fixated on the image. In addition, through BeGaze, the experimenters can highlight specific regions on
a given image, creating areas of interest (AoIs). These areas could either be created using pre-set shapes or by manually tracing the desired area in the structure to be assigned an AOI.

The camera used to track the subjects’ eyes, provided by SMI, was known as the “Red-m camera”, and it was run using the software Iview Red. The camera was active while the participants were taking the experiment on Experiment Center, but was not used during analysis via BeGaze. The camera works by tracking the infrared light reflected from a subject’s pupils. It should be noted that subjects from the baseline visit as well as some subjects from group C were wearing the electrode cap from the EEG experiments. Therefore, at times the camera interpreted an electrode as a pupil, whereupon the experimenter readjusted the camera until it locked properly on the subject’s pupils. The camera was “adjusted” by waving a hand briefly in front of the screen and allowing the camera to re-calibrate.

Additional materials included a Dell monitor and keyboard for the participant to interact with the experiment, and an SMI laptop for the experimenter to track the progress of the experiment as well as how well the camera picked up the participant’s gaze.

**Design:**

Prior to baseline testing, the experimenters chose twenty two images to include in the procedure: seven from the back and limbs section, eight from the thorax abdomen and pelvis section, and seven from the head and neck section. Each of the images was similar to what students saw in the course during their lab dissections. At the end of each
section of the Gross Anatomy course, students take an exam consisting of a written portion and a practical portion, the latter of which is based on what students learn in the anatomy lab. On the practical exam, a particular structure is tagged in a dissection, and students must identify the tagged structure or identify a particular feature related to the structure. The questions in our experiment were generated in a similar fashion, such that a particular structure on a dissection image was tagged. All the questions in the experiment asked subjects to identify a particular structure. The images were chosen such that the participant could identify the structure by looking at nearby attachments or other structures for clues.

For each of the images, the experimenters determined areas of interest that were cognitively salient—that is, areas that provide information about the indicated structure, such a specific attachment or insertion. The researchers for the lab had taken the Gross Anatomy course, and the Principle Investigator is the course director. The cognitively salient areas of interest were therefore determined among the three investigators using their collective anatomy knowledge and experience. In addition, throughout the course, students are taught to look for particular areas when attempting to identify a structure: for example, students are taught to look for the sympathetic ganglion in order to identify a structure as the sympathetic trunk (see Appendix II). During baseline testing, all subjects saw the 22 images mentioned. Following baseline testing, the experimenters added seven obscure images, consisting of animal dissections. The experimenters tagged structures that students had learned in their Anatomy course. However, because the structures were
in animals instead of humans, students would be less inherently familiar with the dissection.

**Procedure**

Subjects completed the task one at a time in room L152, located in the Spivak center. Before the experiment, participants filled out a consent form informing them of the procedure, compensation, and their right to withdraw from the experiment at any point (see Appendix for consent form). The task utilized two monitors: a laptop for the experimenter and a larger monitor for the participant. The gaze tracking camera and chinrest were located in front of the participant’s monitor. The participant rested his or her head on the chinrest such that he or she was 60 cm from the camera, and the camera was able to capture both pupils. Once the participant said he or she was ready, the experimenter administered the experiment via his own monitor.

The first slide consisted of a set of instructions, informing subjects they would be asked to identify a particular structure on a dissection image as indicated by an arrow. Subjects were shown two practice images so that they would grow familiar with the process. Following the practice images but prior to the experiment, the experimenter calibrated and validated the subject’s eye position relative to the camera using the standard SMI procedure. During this procedure, participants viewed a white circle with a red dot in the center of the screen. Participants fixated on a red dot in the center of the computer screen and pressed spacebar once they had focused properly. The shape then moved to all corners of the screen and participants were instructed to maintain their
visual focus. At the end of this procedure, the participants’ degree of accuracy was displayed, where higher numbers represented worse calibration. If participants were more than one degree off, the experimenter repeated the calibration/validation procedure. Participants then viewed a series of dissection image on their monitor and were instructed to identify the tagged structure. Participants could either wait for 60 seconds or press spacebar to advance. On the next slide, participants typed the structure indicated on the previous slide. If subjects could not remember the name of the structure, they were asked to give a descriptive or functional definition. If the subject could not do this, he or she was instructed to type “I don’t know.” For the baseline group, which included all of the participants, the images used were from the Back & Limbs, Thorax Abdomen Pelvis, and Head & Neck sections. When participants re-visited, an obscure session of animal dissections were added. The sections were administered in the above mentioned order (i.e. B&L images were shown first, followed by TAP images, followed by H&N, followed by Obscure), but images within each section were randomized. The experimenter was able to monitor the subject’s gaze tracking pattern throughout the task, and could adjust the camera if necessary. The experiment lasted for approximately 20 minutes.

**MRT Procedure:**

The MRT (Mental Rotation Test) is a task designed to test visuospatial ability by assessing the ability to mentally rotate a shape. In the test, participants must match the correct shape to a “key” shape shown for each trial. Participants may choose from four
shapes for each key, and the correct answer was a rotated version of the key shape (: 
Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-
dimensional spatial visualization. *Perceptual and Motor Skills, 47*(2), 599–604). Prior to 
the experiment, all subjects completed the MRT task, administered on paper. There 
were 20 total trials, and participants were given three minutes to complete as many of the 
first ten as they could, and three more minutes to complete the subsequent ten. 
Participants were allowed a break after the first three minutes. It should be noted that 
participants completed the MRT in a session that was separate from when they completed 
the gaze tracking experiment.

**Variables of Interest generated by BeGaze:**
The variables examined in this study included time spent on a cognitively salient area of 
interest (COGAOIFT) and time spent on the entire image (end time). The data were 
exported to an excel file where they were analyzed. Data were organized based on 
section (B&L, TAP, H&N, O) as well as visit (baseline, visit A, visit B, visit C). Prior to 
analysis, one image each from the Back and Limbs section, Thorax Abdomen Pelvis 
section, and Head and Neck section were removed, since the cognitively salient areas for 
these images were not consistent with the other images.

**Dwell Time/tracking ratio:** The percentage of the participant’s gaze picked up by the 
camera during the time a stimulus appeared was represented as dwell time. Throughout 
the experiment, the infrared light reflected from the participant’s pupils was monitored. 
For the time each stimulus was up, the camera noted the time at which the subject’s gaze
was able to be tracked. When the software lost track of the subject’s gaze, this was accounted for in the variable “dwell time/tracking ratio”. Very few subjects had a tracking ratio of 100 percent, but if the camera ended up picking up less than 80 percent of the subject’s tracking ratio on a given stimulus, the data point for this stimulus was dropped.

**Cognitively Salient Area of Interest fixation time (CogaoiFT):** Each image was assigned one or more cognitively salient areas of interest. BeGaze generated the fixation time on each area of interest generated. For each image, the experimenter calculated the total time spent on each cognitively salient aspect of the image. These were then added, and averaged for each section.

**End Time:** The total amount of time spent on each image. The end time on each image was generated directly from Begaze. The experimenter then averaged the end times based on each section. Thus, there was an average end time for Back and Limbs, Thorax Abdomen Pelvis, Head and Neck, and the Obscure sections.

**Covariates:** In addition to the variables generated through BeGaze, the experimenters monitored other measures of students’ proficiency, including:

1. **Students’ final grades for the course (percent).**
2. **Students’ final practical grade (percent).**
3. **Time spent in the course-based on group (i.e. group A spent less time in the course at revisit than group B at revisit, who spent less time in the course than group C. at revisit).**
4. *Correctness in the experiment*: the experimenters assigned a number to each of the participants’ answers, ranging from 0-2. These numbers represented completely incorrect (0), somewhat correct (1), or completely correct (2). Participants’ answers were considered partially correct if they described the structure or function accurately, but were not able to generate the name. For the purposes of this study, scores of both 0 and 1 were considered incorrect.

**Obscure images:** Another important covariate to consider was each subject’s familiarity with a given section. Thus, groups A and B would be familiar with the B&L section and the TAP section, but not the H&N section. The familiarity would be likely to influence the way subjects viewed a given image; subjects may spend less time on images they know they are not familiar with. Thus, when analyzing a variable, such as CogAOIFT as a function of group regardless of section, the increased familiarity across groups may affect the COGAOIFT. However, with the obscure images, this was not as much of an issue, since each group would share the same amount of familiarity, or lack thereof, with the animal dissections. Thus, change in COAOIFT or End Time for just the obscure images as a function of group would indicate that time spent in the class influenced the way participants looked at an images in general, regardless of familiarity.
**Analysis:**

Once the experimenter collected the data from all four groups (baseline included), the experimenters used several analytical techniques to determine whether learning could be observed using gaze tracking:

---

**CogaoIFt analysis**

For each image, the experimenter calculated the total fixation percentage of the cognitively salient areas. Thus, each subject was organized according to group (A-C), and the experiment was divided into four sections—BL, TAP, HN, O. The experimenter performed a non-parametric ANOVA to compare COGAOIFT across groups (regardless of section), across sections (regardless of group), across just obscure images, and across just familiar images.

---

**End Time Analysis**

As with time spent on cognitively salient areas of interest, the experimenter analyzed End Time as a function of section (Back and Limbs, Thorax Abdomen Pelvis, Head and Neck, Obscure), as well as group (A-C). Analysis was conducted using a nonparametric ANOVA (kruskal-wallis). When including the obscure section in the analysis, the baseline data was not included, since the participants did not see these images during their baseline visit.
The experimenter analyzed the results using the statistical software JMP.

**Correlations:**

1. In order to analyze covariates, we correlated several of the variables mentioned in order to observe whether a relationship existed which may moderate the above analyses. The first correlation concerned MRT score: specifically, did any correlation exist between MRT score and either students’ final grade or final practical grade? The experimenters also parsed the above analyses (CogAOIFT, End Time) based on MRT score. Students were divided into two groups: one group consisted of students with an MRT score of 60 percent and above, while another group consisted of students with an MRT score of 40 percent and below. No student attained an MRT score of between 40 and 60 percent. We performed this analysis in order to assess whether learning progression in a visuospatially demanding task was moderated by students’ innate visuospatial abilities
RESULTS

Cognitively salient AOI analysis:

We first investigated whether groups, regardless of section, spend more time looking at cognitively salient AoIs as a function of time spent in the course (i.e. GROUP). We only compared groups A-C, since the obscure images were included in addition to images from the B&L, TAP, and H&N sections. A chi square revealed that participants spent significantly more time in groups B and C than group A in their fixations on cognitively salient images (COGAOIFT). F(2)=6.98, p<.05. However, there was no significant difference between groups B and C in terms of CogAOIFT (see Figure 1): F(2)=.6, p=0.7. To control for MRT score, the experimenters repeated the previous analyses, but separated the groups into subjects with MRT scores of above 60 and MRT scores below 40. The first analysis was on whether cognitively salient AOI fixation as moderated by group was affected by MRT score. Here, images from all four sections were included. For MRT below 40: a Chi Square revealed that there was a significant difference in CogaoiFT as a function of group: F(2)=5.94, p=.05. Specifically, group C spent significantly more time on the images than group A (p=.02). Group B also appeared to spend more time on the cognitively salient AOIs than group A, but the p value did not quite approach significance (p=.06) (see fig 1x). For MRT above 60: no significant difference was observed in COGAOIFT as a function of group (see fig 1y).

We also analyzed time spent on cognitively salient areas of interests across groups for the obscure images only. Specifically, did participants exhibit significantly different time
spent on cognitively salient images on obscure images only \textit{as a function of group}?

Since the baseline group did not see the obscure images, groups A-C were compared. Although the results indicated that groups B and C appeared to spend more time on cognitively salient images with respect to obscure images (see Figure 2), a Chi square revealed no significant differences in COGAOIFT for obscure images: \(F(2)=5.25, p=.07\).

We then analyzed whether cognitively salient fixation time on obscure images only was moderated by MRT score. For participants with an \textbf{MRT Below 40}: a Chi Square revealed no significant difference, but it appeared that group B spent significantly more COGAOIFT on obscure images than group A: \(F(2)=5.45\ p=.06\). (see fig 2x) For the students with an \textbf{MRT above 60}: no significance was observed. (see fig 2y).

The next analysis controlled for familiarity: specifically, among subjects who are familiar with a given section, will time spent on cognitively salient AoIs differ? Groups A-C were compared for the Back and Limbs section, and groups B and C were compared for the Thorax Abdomen Pelvis section. We were not able to perform a comparison of the Head and Neck section, since only group C was familiar with this. A Chi square test revealed no significant difference in COGAOIFT for the BL section as a function of group (see fig 3a): \(F(2)=2.5, p=.28\). No significant difference was found with respect to COGAOIFT for TAP as a function of section- groups B and C did not differ in their fixation time for TAP images (see fig 3b): \(F(2)=0.9, p=.33\).

Finally, we controlled for MRT on our analysis that had already controlled for familiarity. For the Back and Limbs section, the experimenters analyzed groups A-C, whereas for the TAP section, group A was excluded, since they had not learned the
material when they saw the images. For the **Back and Limbs section (groups A-C):**

**MRT below 40:** A chi square revealed no difference in COGAOIFT for the back and limbs section among any of the groups: $F(2)=2.95, p=.2$ (see fig 3ax). For **MRT above 60:** A chi square revealed no difference in COGAOIFT for the back and limbs section among any of the groups: $F(2)=.49, p=.77$ (see fig 3ay). We then analyzed the **Thorax Abdomen Pelvis section (groups B and C).** For students with an **MRT below 40:** A chi square revealed no difference in COGAOIFT for the back and limbs section among any of the groups: $F(1)=1.28, p=.25$ (see fig 3bx). For students with an **MRT above 60:** A chi square revealed no difference in COGAOIFT for the back and limbs section among any of the groups: $F(1)=.001, p=.99$ (see fig 3by).

Finally, we analyzed whether participants spent significantly different amounts of time on cognitively salient images as a function of section, regardless of their group. A Chi square revealed that participants spent different COGAI fixation times as a function of section: $F(3)=30.86, p<.01$. A Post Hoc revealed that students spent significantly more COGAOIFT on the BL and HN sections compared to the TAP section. In addition, students spent significantly more COGAOIFT on the TAP section compared to the O section. The COGAOIFT between BL and HN was not significantly different. (see fig 1b). The last analysis we completed was the baseline analysis of cognitively salient area fixation as well as end time. A kruskal wallis ANOVA revealed that groups B and C spent significantly more time on cognitively salient features of images than group A (see fig 5): $F(3)=21.9, p<.01$. 

22
**End Time Analysis:**

With respect to total time spent on an image (end time), we asked the following:

Irrespective of section, do groups differ in the amount of time they spent looking at an image *as a function of time spent on the course (I.E. GROUP)*? A chi square test revealed that participants spent significantly more time on images in groups B and C than group A with respect to end time. \( F(2)=15.92, p<.05 \). However, there was no significant difference between groups B and C in terms of end times (see fig 4a).

We then analyzed whether end time varied as a function of section, regardless of group. A Chi square revealed that participants spent different amounts of time on an images *as a function of section*: \( F(3)=22.81, p<.05 \). A Post Hoc revealed that students spent significantly more End Time on the Back and Limbs section than the Head and Neck section (see fig 4b). No other significance was observed.

When we conducted our analysis of end time as a function of group and included the baseline, a non-parametric ANOVA indicated that groups B and C spent significantly more time on images overall compared with groups A and the baseline group (see fig 6)

**Grades and MRT score correlations**

No significant correlations were obtained through correlating students’ MRT scores with either their final grade (\( R=.32, \)) or their practical grade (\( R=.27 \)): see figures A and B.
Figure 1a: Mean percentage of fixation time on cognitively salient areas of interest (CogaoiFT-measured as a percentage) as a function of time spent in the course (GROUP). Participants in groups A, B, and C had recently completed the back and limbs, thorax abdomen pelvis, and head and neck exams respectively. The graph above shows the distribution, represented by points, and the means of each group, represented by the ovoids. The baseline group was not included because participants in this group did not see the obscure images, which were included in this analysis. A chi square revealed that participants spent significantly more time in groups B and C than group A with respect to COGAOIFT.

Figure 1b: Mean percentage of fixation time on cognitively salient areas of interest (CogaoiFT-measured as a percentage) as a function section. The graph above shows the distribution, represented by points, and the means of each group, represented by the ovoids. The baseline group was not included because participants in this group did not see the obscure images, which were included in this analysis. Participants spent different COGAOI fixation times as a function of section: F(3)=30.86, p<.01. A Post Hoc revealed that students spent significantly more COGAOIFT on the BL and HN sections compared to the TAP section.
Figure 2: Mean percentage of fixation time on cognitively salient areas of interest (CogaoIFT - measured as a percentage) as a function of time spent in the course (GROUP) with respect to obscure images only. Participants in groups A, B, and C had recently completed the back and limbs, thorax abdomen pelvis, and head and neck exams respectively. The graph above shows the distribution, represented by points, and the means of each group, represented by the ovoids. The baseline group was not included because participants in this group did not see the obscure images, which were included in this analysis. F(2)=5.25, p=.07.

Figure 3a: Mean percentage of fixation time on cognitively salient areas of interest (CogaoIFT - measured as a percentage) as a function of time spent in the course (GROUP) with respect to back & limbs images only. Participants in groups A, B, and C had recently completed the back and limbs, thorax abdomen pelvis, and head and neck exams respectively. The graph above shows the distribution, represented by points, and the means of each group, represented by the ovoids. The baseline group was not included because participants in this group were not familiar with the back & limbs section.
Figure 3b: Mean percentage of fixation time on cognitively salient areas of interest (CogaoiFT-measured as a percentage) as a function of time spent in the course (GROUP) with respect to thorax abdomen pelvis images only. Participants in groups B, and C had recently completed the thorax abdomen pelvis, and head and neck exams respectively. The graph above shows the distribution, represented by points, and the means of each group, represented by the ovoids. Neither the baseline nor group A were included because they were not familiar with the thorax abdomen pelvis section.

Figure 1x: For participants with an MRT score of 40 or below: mean percentage of fixation time on cognitively salient areas of interest (CogaoiFT-measured as a percentage) as a function of time spent in the course (GROUP). Participants in groups A, B, and C had recently completed the back and limbs, thorax abdomen pelvis, and head and neck exams respectively. The graph above shows the distribution, represented by points, and the means of each group, represented by the ovoids. The baseline group was not included because participants in this group did not see the obscure images, which were included in this analysis. Group C spent significantly more time on the images than group A (p=.02). Group B also appeared to spend more time on the cognitively salient AOIs than group A, but the p value did not quite approach significance (p=.06).
Figure 1y: For participants with an MRT score of 60 or above: mean percentage of fixation time on cognitively salient areas of interest (CogaoiFT-measured as a percentage) as a function of time spent in the course (GROUP). Participants in groups A, B, and C had recently completed the back and limbs, thorax abdomen pelvis, and head and neck exams respectively. The graph above shows the distribution, represented by points, and the means of each group, represented by the ovoids. The baseline group was not included because participants in this group did not see the obscure images, which were included in this analysis.

Figure 2x: For participants with an MRT score below 40: Mean percentage of fixation time on cognitively salient areas of interest (CogaoiFT-measured as a percentage) as a function of time spent in the course (GROUP) with respect to obscure images only. Participants in groups A, B, and C had recently completed the back and limbs, thorax abdomen pelvis, and head and neck exams respectively. The graph above shows the distribution, represented by points, and the means of each group, represented by the ovoids. The baseline group was not included because participants in this group did not see the obscure images.
Figure 2y: For participants with an MRT score above 60: Mean percentage of fixation time on cognitively salient areas of interest (CogaoIFT-measured as a percentage) as a function of time spent in the course (GROUP) with respect to obscure images only. Participants in groups A, B, and C had recently completed the back and limbs, thorax abdomen pelvis, and head and neck exams respectively. The graph above shows the distribution, represented by points, and the means of each group, represented by the ovoids. The baseline group was not included because participants in this group did not see the obscure images.

Figure 3ax –for participants with an MRT score below 40: Mean percentage of fixation time on cognitively salient areas of interest (CogaoIFT-measured as a percentage) as a function of time spent in the course (GROUP) with respect to back & limbs images only. Participants in groups A, B, and C had recently completed the back and limbs, thorax abdomen pelvis, and head and neck exams respectively. The graph above shows the distribution, represented by points, and the means of each group, represented by the ovoids. The baseline group was not included because participants in this group were not familiar with the back & limbs sections. Despite the apparent tendency towards significance, the difference in COGAOIFT on obscure images as a function of section never approached significance.
Figure 3ay – for participants with an MRT score above 60. Mean percentage of fixation time on cognitively salient areas of interest (CogaoiFT-measured as a percentage) as a function of time spent in the course (GROUP) with respect to back & limbs images **only**. Participants in groups A, B, and C had recently completed the back and limbs, thorax abdomen pelvis, and head and neck exams respectively. The graph above shows the distribution, represented by points, and the means of each group, represented by the ovoids. The baseline group was not included because participants in this group were not familiar with the back & limbs section. No significant difference was observed between groups.

Figure 3bx – for participants with an MRT score below 40: Mean percentage of fixation time on cognitively salient areas of interest (CogaoiFT-measured as a percentage) as a function of time spent in the course (GROUP) with respect to back & limbs images **only**. Participants in groups A, B, and C had recently completed the back and limbs, thorax abdomen pelvis, and head and neck exams respectively. The graph above shows the distribution, represented by points, and the means of each group, represented by the ovoids. The baseline group was not included because participants in this group were not familiar with the back & limbs section. No significant difference was observed between groups.
Figure 3by –for participants with an MRT score above 60. Mean percentage of fixation time on cognitively salient areas of interest (CogaoIT-measured as a percentage) as a function of time spent in the course (GROUP) with respect to back & limbs images only. Participants in groups A, B, and C had recently completed the back and limbs, thorax abdomen pelvis, and head and neck exams respectively. The graph above shows the distribution, represented by points, and the means of each group, represented by the ovoids. The baseline group was not included because participants in this group were not familiar with the back & limbs section. No significant difference was observed between groups.

Figure 4a: Total time spend on each image (End Time-measured in milliseconds) as a function of group (A, B, or C). Participants in groups A, B, and C had recently completed the back and limbs, thorax abdomen pelvis, and head and neck exams respectively. The graph above shows the distribution, represented by points, and the means of each group, represented by the ovoids. The baseline group was not included because participants in this group did not see the obscure images, which were included in this analysis. Results indicated that group C spent significantly more time overall on the images than group A.
Figure 4b: Total time spend on each image (End Time-measured in milliseconds) as a function of section (BL, HN, O, or TAP). The graph above shows the distribution, represented by points, and the means of each group, represented by the ovoids. The baseline group was not included because participants in this group did not see the obscure images, which were included in this analysis.

Figure 5: Comparison of fixation on cognitively salient areas of interest with the baseline group included (base). The obscure images were excluded from this analysis because the baseline group had not seen these images. The points on the graph represent the distribution and the ovoids represent the means. A post hoc revealed that students in groups B and C spent significantly more time fixating on cognitively salient areas of interest compared to the baseline group or group A.
Figure 6: comparison of end time as a function of group, with the baseline included. The obscure images were excluded from the analysis since the baseline group did not see these images. A post hoc revealed that students in groups B and C spent significantly more time on images overall compared to groups A and C.

Figure a: Correlation of students’ MRT scores with their final grade. No significance was observed.
Figure B: Correlation of students’ MRT scores with their final practical grades. No significance was observed.
DISCUSSION

Our experiment was designed to assess whether it was possible to measure learning progression in a visuospatially demanding course by studying gaze tracking variables. The proposed hypothesis was that students would spend a greater relative percentage of time on cognitively salient areas of an image as they spent more time in the class. We added obscure images in order to see whether this change in cognitively salient fixation time occurred regardless of whether students were familiar with the material. Thus we assessed whether the students acquired a generalized ability to pick out cognitively salient features of an image as a function of time spent in the course, even if they had never seen the image before. We had also hypothesized an effect of end time across groups such that as groups spent more time in the course, they would spend more total time on each image/stimulus.

Overall, the results obtained were ambiguous, and warrant further study in order to arrive at a clear conclusion. With respect to cognitively salient AOI fixation time, students spent more time on the images overall as a function of time in the class: students who spent more time in the class (groups B and C) showed overall greater cognitively salient AOI fixation times than the earlier groups or baseline. As mentioned in the methods, though the stimuli were randomized within their section, the sections themselves were discreet. Therefore, it may have been possible that students in group A, knowing that they had not learned images from Thorax Abdomen Pelvis or Head and Neck, did not bother to spend time trying to identify the stimuli. Familiarity was
controlled for by testing groups A-C on the Back and Limbs images only and testing groups B-C on the TAP images only. In the case of the Back and Limbs images, it appeared that groups B and C spent significantly more time on cognitively salient features than group A. However, it should be noted that there was no significant difference between groups B and C for time spent on cognitively salient features of the Thorax Abdomen Pelvis section. When controlling for MRT, the significance observed for the back and limbs section disappeared, but this may have been due to fewer subjects, since the values still tended towards significance.

The fact that groups B and C no longer differed when coding for familiarity may indicate that students likely spent less time on images they knew they had not yet learned. Of particular interest is the fact that for the back and limbs section, groups B and C spent significantly more time on cognitively salient features than group A. One possible reason for this may be the fact that group A had just learned the back and limbs section, and therefore may have stored the image in more recent memory than subsequent groups. Therefore, this group would not need to spend as much time looking at the cognitively salient areas of interest, since they had memorized the image as a whole. However, this theory is called into question given that for the TAP section, groups B and C were not significantly different, even though the section should be more recently familiar for group B.

Familiarity was also controlled for by including the obscure images. Although the structures that were tagged were structures students had learned, the form of the structures differed since they were in animals. Thus, students would be more forced to
rely on nearby clues for these images than for images they could have committed to memory. Although the results obtained were not significant, there was a trend such that groups B and C appeared to spend more time on cognitively salient features of the image than group A. It is possible that there is a transition in the course, between groups A and B, where students show a notable change. It may be possible that leading up to and immediately following the first exam, students are still getting used to the demands of the course. However, leading up to and following the second exam, students may have a better grasp of how to approach the material.

The results for end time as a function of group indicated that subsequent groups tended to spend more total time on the images than earlier groups. However, the data only approached significance when comparing group C to group A. This was likely due to the fact that subjects from earlier groups did not spend as much time on sections they knew they had not learned.

There did not appear to be a main effect of end times as a function of section, meaning the groups did not differ in the amount of time they spent on BL, TAP, HN, or O sections of the experiment. This is an encouraging result, since it indicates that in our experiment, no one particular section lent itself to more time. The same was not true for cognitively salient AoI fixation time; in particular, it appeared that students spent significantly more time on cognitively salient images on the Back and Limbs and Head and Neck sections compared to the Thorax Abdomen Pelvis and Obscure sections. The experimenters designed the cognitively salient areas of interest based on their anatomy knowledge, but it is possible that some sections contain more cognitively salient features
than others. It is also possible that the relative area of these cognitively salient features varied across sections. It is worth noting that the obscure section contained the least amount of fixation time on cognitively salient features, which is counterintuitive given that these are not familiar images. One possibility is that the cognitively salient areas were not as well defined on the obscure images. In addition, students were told that the images were from animals, which may have affected the way they fixated on various features. In future experiments which include obscure images, it may be more conducive to randomize this section within the experiment.

**Future Directions**

The design and results of this experiment provide insight as to how to conduct future studies that assess learning through gaze tracking. A particular issue that was not explicitly addressed in this experiment may be to find out at what point there is a noticeable transition in the learning progression of students. In the future, it may be worthwhile to break students up into more groups, rather than only testing them after an exam. It may be worth it, for example, to test at a point halfway between the first exam and the second one. A possible problem with this would be that there would be fewer students per group, so any trends may be stifled due to the small N. An alternative would be to break students up into 3 groups, but test them at the halfway points rather than directly after the exam. This would maintain the number of students per group, and it would also make students less likely to be overly familiar with the section they had just learned.
It may also be worth further investigating whether the type of image lends itself to different fixation techniques. In addition to the section to which the images belong, one could classify the image based on whether it is a cross section, osteology section, or a dissection. It is possible that students learn slightly different techniques as a function of the type of image—perhaps they are more likely to commit cross sections to memory, since in the Gross Anatomy course, they are responsible for knowing particular sections.

Gaze tracking is still a relatively nascent field, so further experiments should be performed to arrive at the variables most conducive to measuring learning. Previous research has posited that the difference in gaze patterns between novices and experts lies in fixation on areas that are not necessarily visually salient, but information yielding (e.g. Wood 1999, Matsumoto 2011). It is worth noting, however, that “experts” were defined as those who had been in their field for several years (e.g. Nodine 2009). Therefore, a more sensitive measure may be required to measure students’ progression across a semester. One variable worth measuring may be the time at which students first fixate on a cognitively salient area of interest. Students may have spent a significant amount of time on a cognitively salient feature, but may have done so coincidentally or due to its position in the image. Research has shown that experts also tend to fixate on the cognitively salient areas of interest faster than novices (Matsumoto 2011, Myles-Worsley 1988, Parkhurst et al 2003). Therefore, it may be worthwhile to create an aggregate variable to more thoroughly assess expertise (e.g. time spent on cognitively salient area of interest divided by time it took to fixate on the area).
APPENDIX I

Examples of images from the Back and Limbs section, Thorax Abdomen Pelvis Section, and Head and Neck sections respectively. Participants were asked to identify the structure indicated by the arrow. The AoIs are not indicated.

*Structure: Median Nerve
*Structure: Vagus Nerve.
*Structure: Sternocleidomastoid
REFERENCES

1. A Archambault, C O'Donnell, PG Schyns (1999) Blind to object changes: When learning the same object at different levels of categorization modifies its perception Psychological Science


CURRICULUM VITAE

ARJUN IYER

Email: aiyer1@bu.edu, alexiyer12@gmail.com

Born 1988

Work Experience

Research Assistant

SUNY at Stony Brook, Neurology Department, June 2011-June 2012
Assist Dr. Thomas Preston with research involving prospective memory
Duties included statistical input, data analysis, and aiding in writing the abstract
Second author on a 2 posters. One entitled “Neurocognitive Correlates of Prospective Memory in Clinically Referred Children”. The other entitled “Psychosocial and Behavioral Correlates of Prospective Memory in Clinically Referred Childrenj.”
  o Principle Investigator: Dr. Thomas Preston-Stony Brook Neuroscience Dep’t.

Activities Coordinator

St. Mary's Hospital, January 2010- May 2010
Worked with adolescents with conduct disorder and ADHD-hyperactive type
Coordinated activities that encouraged cooperation within groups

Independent Researcher

Antioch University, September 2009- December 2009
Interviewed and shadowed Tibetan doctors in Dharmasala, India
Completed 24-page paper on the various diagnostic and treatment aspects of Tibetan medicine

Group Co-Leader

Child and Family Wellness Center, June 2008-August 2008
Constructed and led group activities for children and adolescents with psychological disorders including depression and ADHD
Coordinated activities including yoga, percussion, art, theater, and teaching communication skills
Education (transcripts available on request)
Bates College – Lewiston, ME
• Bachelor of Arts in Psychology, May 2011
• Dean's list September 2010 – May 2011 (Senior year GPA: 3.74)
Boston University-Boston, MA
• Master of Arts in Neuroanatomy and Biology, May 2014
  ▪ Completed a Graduate Thesis.
  • Principal investigator: Dr. Ann Zumwalt
  • Second author on a paper she published where I helped construct and analyze variables, entitled “Gaze Patterns of Gross Anatomy Students Change with Classroom Learning.”
    o Published in Anatomical Sciences Education.