A Wizard Hat for the Brain: Predicting Long-Term Memory Retention Using Electroencephalography

Noah Libby
Bard College, nl8800@bard.edu

A Wizard Hat for the Brain: Predicting Long-Term Memory Retention

Using Electroencephalography

Senior Project submitted to

The Division of Science, Mathematics, and Computing

of Bard College

by

Noah W. Libby

Annandale-on-Hudson, New York

May, 2019

Dedication

To my brother, Jacob, for being the best damn best man a best man’s ever been.

Acknowledgments

With a full heart, I would like to thank Justin Hulbert for advising me in research, life, and British metal detectorist television shows. Your guidance and leadership have inspired my passion to decode the brain and your friendship has made Preston, and the world outside it, far more than just live-able. I would also like to thank Tom Hutcheon for introducing me to the beauty of Psychology during my first year at Bard, and for sharing your excitement about cognition and statistics with me for the next few years after that.

To Zall, for designing experiments with me and always being down to toss. To Jan, for your everlasting love in the form of nightly hugs. To the Aggressive Apple Snatchers for kidnapping me four times. To the Memory Dynamics Lab for helping me grow as a scientist. To team supreme for being my boys, forever and always. To Jackson, for being my sock buddy and to Meagan for being my adventure buddy, you have shown me real love. To my family for loving me and fortifying me with food and prayers even when I didn't ask for them — I would not be who I am without all of you. And to God, for keeping me grounded and helping me always catch my breath.

Table of Contents

"*The name of the author is the first to go*
followed obediently by the title, the plot,
the heartbreaking conclusion, the entire novel
which suddenly becomes one you have never read, never even heard of,

as if, one by one, the memories you used to harbor
decided to retire to the southern hemisphere of the brain,
to a little fishing village where there are no phones."

Billy Collins, *Forgetfulness*

Abstract

Learning is a ubiquitous process that transforms novel information and events into stored memory representations that can later be accessed. As a learner acquires new information, any feature of a memory that is shared with other memories may produce some level of retrieval-competition, making accurate recall more difficult. One of the most effective ways to reduce this competition and create distinct representations for potentially confusable memories is to practice retrieving all of the information through self-testing with feedback. As a person tests themself, competition between easily-confusable memories (e.g. memories that share similar visual or semantic features) decreases and memory representations for unique items are made more distinct. Using a portable, consumer-grade electroencephalography (EEG) device, I attempted to harness competition levels in the brain by training a machine learning classifier to predict long-term retention of novel associations. Specifically, I compare the accuracy of two logistic regression classifiers: one trained using existing category-word pairings (as has been done previously in the literature), and one trained using new episodic image-name associations developed to more closely model memory competition. I predicted that the newly developed classifier would be able to more accurately predict long-term retention. Further refinements to the predictive model and its applications are discussed.

Introduction

Section 1: Learning and Studying

Learning. A young child learns how to build a campfire that burns for hours by trying different ways of setting up the logs. A baby-boomer learns to code in Python using an online course with weekly project deadlines. A professor learns which teaching methods were most effective as she reads student feedback from the semester. Two strangers learn to have more empathy by taking time to listen to each others' stories. For all of these people, though using distinct strategies and working in very different domains, the end result is the same: They have each learned something new.

Learning is a ubiquitous process that preserves experience and orients behavior. It occurs under various circumstances, is driven by unique motivations, and is the basis for the formation of memories. Learning creates memories for specific events, skills, and knowledge that can then later be accessed when those experiences again become relevant. This allows learners to transcend the linear path of time by accessing their account of the past and using it to direct their actions in the present. By attempting to recall the representation of an experience held in memory, learners can guide their own behavior. As such, a mother will remember how to do her job when she goes into work every day, an athlete will remember the best way to guard a specific defender whilst playing ultimate frisbee, and a student will remember the parts of a neuron for their in-class quiz.

In an ever-changing world, there is an overwhelming multitude of information that can be learned, and sorting through to find what is relevant is a challenge that all learners face. Even more, learners must aim to acquire this information in the most efficient way possible, so as to preserve cognitive resources and yield the greatest results in terms of memory retention.

Though an immeasurably powerful cognitive development, the process of learning and later remembering is imperfect. Memory for both experiences and learned material will often be subject to the fate of being forgotten. While forgetting can result simply from the passing of time (Ebbinghaus, 1885, 1913), there are many forces that influence memory retention. Sources of a memory can be misattributed, the context of the present situation can bias how the past is recalled, and sometimes people just can't remember where they left their keys (Schacter, 1999). And still, people continue to engage in the cyclical process of learning by recognizing the fallibility of their own memory and adjusting their subsequent behavior. The person who forgot their keys will make a mental note to remind themselves about it next time, and the student who failed their test will further space out their study sessions for the next quiz.

While undeniable that learning occurs not only in the classroom, the learning experience of students and how it can be enhanced will be the main focus of this manuscript. The population of students at any given time in elementary to post-secondary schooling, all of whom are essentially full-time dedicated learners, demonstrates a prodigious body of people that would benefit directly from innovative education techniques. Understanding the dynamics of effective learning practices, and how they can be optimized, will allow students and the greater community to be become better learners.

Study Strategies. Throughout their education, students are tasked with learning copious amounts of information from a variety of disciplines, and they are required to do so in little time. With this onslaught of knowledge to be acquired, students tend to fall into patterns for how they learn. These patterns may encompass a number of strategies, some of which are only effective for achieving short-term goals like passing an exam, while others encourage a deeper encoding that creates an environment for more robust learning.

Having both depth and breadth of understanding becomes requisite as students advance and specialize in their studies; being able to learn quickly and retain that learning is thus a valuable skill to develop. Yet, this skill is not useful in the long-term if the information is merely memorized in the context in which it was initially learned. Comprehending information, extracting critical components, and transferring that knowledge to various new settings encourages a deeper learning and understanding of the content that is irrespective of the context in which it was learned.

Stress and anxiety pile up as students spend hours studying (or not studying) for midterm and final exams that will likely determine a significant portion of their course grade. In this common scenario, several factors are at play. Leading up to the exam itself, students are tasked with the responsibility to study all of the material that was discussed in class, assigned for independent readings, and completed during assignments. Of the many modes of learning, students often rely on tactics such as spacing out their study time (or the opposite: cramming), rereading, making outlines, highlighting important passages, recopying notes from memory, and self-testing. However, some of these study habits and modes of learning prove to be more effective than others (Rodriguez et al., 2018).

Putnam, Sungkhasettee, and Roediger (2016) outline methods drawn from cognitive psychology for college students to optimize their learning. One of their suggestions is to prioritize self-testing, or recall of learned material, when studying for a class. This is supported by a growing body of literature on the so-called "testing effect," which has revealed that actively testing learned information yields better long-term memory when compared to rote memorization and classical re-reading habits upon which students often rely (Rowland, 2014). Though it may be tempting and seem easier to simply read through notes or highlight important passages, self-testing produces the greatest learning outcomes when compared to other strategies (Rodriguez et al., 2018). Students would thus benefit greatly, in terms of both final grade and long-term retention of material, if they choose to practice testing themselves as they study. This holds true across various types of studied material and has proven to be effective in actual classrooms, both in-person and online (Broek et al., 2016).

A seminal study on the testing effect outlined the benefit of retrieval-practice in the context of developing effective study habits that yield long-lasting learning. Karpicke and Roediger (2008) were interested in the role that repeated-testing had on long-term memory recall. Participants were divided into four groups and each group was given eight periods of interleaved study and test trials to learn foreign (Swahili/English) word pairs that were unfamiliar to them. One group completed a typical set of study and test periods during which they studied and were tested on all 40 word pairs in each period. A second group was always tested on all word pairs, but dropped items from study periods once they had been correctly recalled. A third group always studied all word pairs, but dropped items from test periods once they had been correctly recalled. And a final group dropped words from both study and test periods after they had been correctly recalled. In the first two periods, all participants studied and were tested on all 40 Swahili/English word pairs and thereafter, the amount of word pairs presented in each period varied across conditions.

The researchers found that participants who were repeatedly tested on all of the information (i.e. those in the conditions that did not drop word pairs from test periods) recalled a greater proportion of word pairs on a final test one week after learning occurred. Importantly, continuing to be tested on all word pairs yielded greater long-term memory regardless of whether or not the words were dropped from the study periods. In contrast to common belief, testing serves a purpose beyond simply assessing the state of knowledge and whether or not something has been sufficiently learned. Testing serves as an active and effective learning tool that produces better long-term memory compared to re-studying.

The delay between acquisition and the final test reflects real-world learning; information will be learned when it is first relevant at an initial point in time, and then memory of that learning will be assessed when the information is once again relevant at a later point in time. While the learner may feel confident in their ability to remember the word pairs immediately after studying them, this judgment of learning may be reflecting an awareness of short-term retention as opposed to long-term, which relies on deeper encoding. This convolution of judging short-term versus long-term retention complicates a learner's ability to accurately assess their learning. In the same study, researchers found evidence that participants indeed had difficulty judging their learning. Participants were unilaterally unaware of the strong benefit of testing in regards to their own learning. When asking participants how many words they believed that they could recall one-week post-learning, there was no significant difference between group predictions.

Despite the advantage of using self-testing as a learning method, students are generally unaware of this benefit and use it to simply assess whether they have learned. This is still a useful metric, but the value of self-testing as a study method in-and-of itself is often poorly understood by students.

Judgments of Learning. Of utmost importance when studying is knowing how to best allocate time. Textbooks and articles can be read and reread indefinitely, but finding the optimal amount of time to study specific material is valuable for a number of practical reasons. First and foremost, learning can be a time-consuming process with no apparent endpoint. Understanding when something has been learned to the point of when it can be recalled from memory later on will allow for reallocation of study time. On a related point, undergraduate students are often enrolled in multiple courses that each require a substantial amount of time and attention in order to be prepared for exams and papers. Knowing when information has been learned for one course will free up valuable time for students who need to balance learning in several courses. Finally, learning can be an exhausting process. Paying attention to study material and learning to the best of one's ability uses limited and valuable cognitive resources. In addition to the fact that students are pressed for time, gaining better insight into one's own mastery of material encourages better learning habits.

People often engage in metacognition: the ability to reflect upon and assess one's own cognitive processes. Students tend to tap into their metacognition when they learn, using it as a way to gauge how well they have learned something. Measuring one's own knowledge, through self-testing, writing notes from memory, or using other evaluation strategies, eases anxiety about whether the information has been learned, helps to further strengthen memory of the information, and highlights areas requiring further improvement. People have some subjective awareness of their abilities, but these are not always complete and accurate representations.

A current limitation that students face while studying is their own imperfect metacognition. This is evident on two fronts. First, students tend to be unaware of the most advantageous study strategies. In a survey of study behaviors in students' natural learning environments, Karpicke, Butler, and Roediger (2009) found further evidence that students do not regularly test themselves as a form of studying. Despite the overwhelming evidence that self-testing is an extremely useful and effective study strategy, students seem to be unaware of its benefits and do not use it. Self-testing may be avoided by students for whom testing is only ever experienced in a stressful context. Second, students tend to generally overestimate how much information they will remember, regardless of whether they engage in self-testing strategies, and they often experience "illusions of competence" during studying that impair long-term memory (Karpicke & Roediger, 2008; Karpicke, Butler, & Roediger, 2009). Metacognition plays against the user in this way: one is both the learner and the one assessing the learner's capabilities. Therefore, any error in judgment can yield extremely negative effects. The learner can convince themselves that, while studying all the atomic weights on the periodic table of elements, they have learned all of them after getting most of them correct. Though it may be reasonable and even valuable to recognize progress in learning, there may be academic suffering if a student's perception of their learning does not align with their actual ability to later access and utilize the information from memory.

Section 2: Predicting the Future

Combining Learning and EEG to Predict Long-Term Memory. Imagine an undergraduate student taking Introduction to Psychological Science. He spends hours in the library studying for the cumulative final exam that looms over him. He worries that he has not studied everything, that he does not know all of the information well enough, and that he will fail the exam and therefore the entire course. His ability to judge whether he has learned everything relies on his own metacognition, which may not be optimal. Perhaps he will miss something or will decide he knows something before he fully understands it, and his final grade will reflect this and suffer as a result. This student, and many like him, would benefit greatly from a more objective judgment of learning — one that relies on a more direct measure of knowledge, giving better predictions about the long-term memorability of learned material.

Such a system might be carried out in the following manner. As a person is studying a set of information, neural signals will be recorded. These signals provide a clearer window into the brain's cognitive state, which is only partially accessible to conscious awareness otherwise. After a predetermined amount of time, the studied information will be tested. At this point, the data will be decoded and analyzed to find consistent patterns that emerge in specific contexts. Patterns that are present when information is later remembered are of particular interest, and finding these patterns will allow for predictions to be made about the long-term fate of specific memories.

Neural signals associated with the encoding of information tend to differ depending on whether the information is later remembered or forgotten, yielding what has been termed a difference due to memory (Dm) (Sanquist, Rohrbaugh, Syndulko, & Lindsley, 1980). Investigating Event-Related Potentials (ERPs) during incidental learning, Paller, Kutas, and Mayes (1987) exposed participants to 300 words while completing a processing task. For each word, participants answered either a semantically dependent question or a non-semantically dependent question. Semantically dependent questions (i.e., "Is it living?" or "Is it edible?") required participants to process the meaning of the word, while non-semantically dependent questions (i.e. “Are there two vowels?” or “Are the first and last letters in alphabetical order?”) required the participant to process only the physical composition of the word. Words were thus differentially encoded as a result of the processing task.

Recognition and recall memory for the incidental encoding of those words were later tested. Researchers found greater positive ERPs over left and right parietal electrodes in the late positive complex (LPC; 400 to 800 ms post stimulus) for items that were later remembered compared to those that were later forgotten. Interestingly, there was a significant interaction between processing task and whether the word was later recognized. LPCs for words that were encoded while being asked semantically dependent questions were even greater than for words encoded while being asked non-semantically dependent questions. The Dm effect observed in this study shows promising methodology for determining a measure of successful encoding using ERP. However, the underlying principles that guide this effect are still unclear. Though there was a greater effect for words processed semantically in Paller, Kutas, and Mayes (1987), there is still a general Dm effect for all words that were later remembered, which requires further dissemination.

Building from research on the Dm effect and the predictive power of Electroencephalography (EEG), Noh, Herzmann, Curran, and de Sa (2014) published further evidence for the ability to predict memory using single-trial EEG data. Participants completed a visual old/new judgment task for a set of images of cars and birds. Participants indicated whether they could recall the image from the learning phase, or they rated their level of familiarity with recognizing it if they had no clear recollection. EEG was recorded across three different time periods: before-stimulus (-300 to 0 ms), early during-stimulus (400 to 800 ms), and late during-stimulus (1000 to 1400 ms). The authors found that signals obtained in the pre- and during-stimulus periods were sufficient to predict the dynamics of long-term memory performance. Specifically, pre-stimulus activity in the high-beta (power in the range of 19-30 Hz) and low-gamma bands (12-19 Hz), as well as during-stimulus activity in the alpha band, was able to differentiate between recognition and recall memory. This evidence supports the notion that stimulus-specific information (in this case, defining features of each car or bird), which is critical for accurate memory recall, is encoded before contextual information (i.e., visual clues that suggest the category of the image), which aids in recognition memory tasks. The authors go so far as to suggest that this methodology can be used to highlight optimal times for learning, given a current brain state, and that targeting those times at encoding and recall may be used to improve memory capabilities.

Another line of research designed to predict subsequent memory performance behaviorally and electrophysiologically explores the underlying reason for the elevated success of the testing effect over other study strategies. Ultimately, self-testing is thought to be so effective because it encourages the strengthening of semantic networks. Often in the form of a cue-target word pair, participants are given the cue and asked to produce the target. Actively attempting to recall the associated target makes memory representations for the association more established. In other words, by trying to complete the association, related information becomes more conceptualized and the memory representation is made clearer. The exact mechanisms that account for such strengthened memory associations remain contentious between two main theories: the elaborative retrieval hypothesis and the search-set restriction account (Broek, 2016).

The elaborative retrieval hypothesis maintains that improved recall attained from repeated testing is a result of an increasing amount of routes to target information. By searching for the correct answer, the semantic network of related information is expanded. Through these new or strengthened connections, people are able to use many different retrieval routes to better access the correct answer (Carpenter & DeLosh, 2006). However, the elaborative retrieval hypothesis is inconsistent with the fan effect — the notion that the ability to recognize words depends on how many items are linked to that word (Anderson & Reder, 1999). If there are more items linked to a given word, the retrieval will be more difficult as one has to search through and suppress or restrain competing alternative responses.

The search-set restriction account, on the other hand, holds that memory competition is at the heart of the testing effect. When cue-target associations are repeatedly tested by presenting the cue, the target response is selectively strengthened while incorrect responses are weakened. For example, repeatedly practicing the retrieval of "VODKA" given the cue, "ALCOHOL-V" will strengthen the association between the cue and the response, "VODKA," while impairing the accessibility of other potential responses, such as "VERMOUTH". These results are consistent with the literature on retrieval-induced forgetting, which shows that selectively retrieving a target response to a cue makes related responses less accessible (Anderson, Bjork, & Bjork, 1994; Murayama, Miyatsu, Buchli, & Storm, 2014). When associations are repeatedly practiced via testing, memory representations are able to become more established as competing memories are made more distinguishable. As a result, distinguished memories develop long-term stability. This work can help refine the findings from the Dm effects in order to create better predictors of long-term memory.

It is important here to note that memory competition can arise in a variety of contexts. Any feature of a memory that is shared with other similar memories may produce some level of competition. Imageable memories might find themselves in competition with one another if they have similar visual features, or words in a list might compete with one another if they look, sound, or are spelled similarly. In a category-exemplar task, cueing the participant with the category means that participants have to sift through a larger set of competing memories, compared to when they must retrieve the category when cued with an exemplar. Participants are thus better at recalling "ALCOHOL" when prompted with "A-VODKA" because there is only one potential category associated with the exemplar, "VODKA."

Similarly, when learning a new language, competition between the visual, phonic, and semantic features of words makes it difficult to retrieve the correct translation. Learners are given the difficult task of distinguishing newly learned information from other new and previously acquired knowledge. Navigating through a large network of memories to find the relevant and correct target is a kind of competition that arises frequently in real-world learning situations in which a lot of related information has to be learned and integrated with existing frameworks of knowledge. Each type of competition may help to further differentiate individual items.

Rafidi, Hulbert, Brooks, and Norman (2018) used the notion that repeated testing diminishes memory competition in order to develop a machine learning classifier that predicts long-term retention. Using EEG, their work found neural oscillatory signatures of memory competition that have been shown to predict subsequent performance on associative memory tasks. Rafidi et al. (2018) trained a machine learning classifier to distinguish "high" versus "low" levels of retrieval competition using EEG recordings while participants learned Swahili/English word pairs. The experiment took place across three sessions. In Session One, participants studied a set of English category-exemplar pairs and then completed four blocks of high- and low-competition cued recall. In high-competition blocks, participants had to retrieve the exemplar given the category and the first two letters of the exemplar. In low-competition blocks, participants had to retrieve the category given the exemplar and first two letters of the category. EEG data from this session were used to train a logistic regression classifier to differentiate between signatures for low- and high-competition. In Session Two, participants completed eight blocks of interleaved study and retrieval-practice. Participants were presented with the Swahili word and its English translation during study blocks. During test blocks, participants waited two seconds and then typed the English translation when presented with the Swahili word. Finally, one week after the initial encoding period in Session Two, participants returned to the lab and were tested on their memory for Swahili/English word pairs; given the Swahili word, participants had to recall the English translation.

Using the classifier trained in Session One, the authors inputted EEG data from Session Two as a testing dataset. Based on what the classifier learned when trained during Session One, it made predictions about competition levels in the testing dataset. As levels of recorded competition decreased, the classifier accurately predicted better long-term recall of the learned vocabulary after a week-long delay; 64% classification decoding accuracy at $p < 0.0001$.

While Rafidi et al. (2018) present evidence that suggests memory competition is a prominent factor that influences the long-term memory benefits seen in the testing effect, several concerns remain regarding the implementation and operationalization of competition. First, the

EEG training phase of the Rafidi et al. (2018) experiment does not account for certain uncontrolled variance, such as differences in type of competition between training and test phases, and the inherent pre-existing semantic associations between category-exemplar pairs. In the classifier training phase, pre-existing category-exemplar pairs are leveraged to create high and low levels of competition by either cuing participants with a category and requiring retrieval of the exemplar, or vice-versa. This is not equivalent to the kind of search-set competition that is present as participants try to recall the proper English definition given the Swahili translation. Competition during this Swahili/English task is likely greater because it occurs at the time of encoding. It is also noteworthy to mention that the category-exemplar associations may already have pre-existing connotations for different participants, as some have more experience with them in the real world more than others. Using the example of alcoholic beverages, certain people may be more knowledgable of alcohol than others, thereby making the degree to which information is known inconsistent between participants. These variables could produce noisy data that makes accurately classifying relevant EEG characteristics more difficult. The proposed experiment trains a classifier that reflects a more generalized learning experience in which complex associations are formed in the laboratory, thus affording more control and a greater likelihood of accurately classifying a kind of memory competition that is relevant for real-world learning by aiming to classify the dynamics of competition at the time of encoding.

While competition may be a strong theoretical construct to measure learning, there may be more effective ways of isolating it operationally. Hulbert and Norman (2015) explore memory competition using a different methodology, fMRI BOLD responses, to track changes in hippocampal pattern similarity as people learn to distinguish competing memories. Using a neural network model inspired by Norman, Newman, Detre, and Polyn (2006), the authors found that as participants completed interleaved study and retrieval-practice trials, hippocampal representations for individual but related items became more distinct. That is, the activity occurring in the hippocampus, as recorded through the fMRI scanner, was different for distinguished items. At the same time, cued-recall memory for items that received retrieval-practice was significantly greater than items that did not receive retrieval-practice at a final test. This yields further evidence that the memorability of a target item is, at least in part, due to the differentiation of the target item from other related items, and that this extinguishing of competition occurs on the neural level. While the fMRI methodology used in this study may be less practical for everyday learners who may eventually use similar technology to augment learning practices, the materials and elements of the procedure could be useful for isolating competition between memories that exist in a large network of associations.

Memory competition seems to be a fruitful and practical mode through which learning can be understood on both the behavioral and neural levels. As a person learns the difference between the Swahili words *mashua* (boat) and *maziwa* (milk), they begin to misinterpret them less often, despite their sharing similar spellings. With practice, the intricacies of the individual words become more apparent and the differences more stark. At the same time, the learner's hippocampus is differentiating the once apparently similar words. By making each Swahili word a distinct neural representation, the words have a better likelihood of later being correctly remembered.

Method

The method outlined for this experiment was approved by the Bard College Institutional Review Board (IRB; Appendix 6). Over the course of three sessions, participants completed two classifier training procedures (*Episodic Animal Family Classifier* and *Semantic Category-Word Classifier*) and two classifier testing procedures (*Show and Tell Task* and *Swahili/English Task*).

Participants

Eight participants between the ages of 19 and 30 (M = 22, SD = 3.54) completed the informed consent process and participated in all three sessions of the experiment. Five participants identified as male, two identified as female, and one identified as non-binary. Participants were run between February 27, 2019 and April 15, 2019, with half of the participants completely run before the spring intersession and the other half completely run after. Full counterbalancing was achieved across the final sample of eight participants.

Instruments and Data Method

MATLAB Scripts. In an effort to encourage and support the Open Science movement and create transparent research, all of the code designed, used, and written for this experiment has been uploaded to the Internet for public use and can be found freely available in an online repository hosted on GitHub[1].

The scripts for the *Semantic Category-Word Classifier* and the *Swahili/English Task* were the same materials used in Rafidi et al. (2018) and were written in MATLAB R2013a using Psychtoolbox-3. The scripts for the *Episodic Animal Family Classifier* and the *Show and Tell Task* were also written in MATLAB using Psychtoolbox-3. Additional scripts for preprocessing

[1] https://github.com/noahlibby17/compN-Senior-Thesis

and analysis used the MATLAB Signal Processing Toolbox and Fieldtrip, an open source MATLAB toolbox for neurophysiological data interpretation and analysis (Oostenveld, 2011).

EEG Acquisition and Event Marking. Electroencephalographic (EEG) data were collected using the Emotiv EPOC+, a prosumer, portable EEG headset (Figure 1). Emotiv headsets provide scalable and affordable solutions for EEG-based research, and have been the subject of a growing body of the published literature (see Ekandem, Davis, Alvarez, James & Gilbert, 2012; Badcock et al., 2013; Badcock et al., 2015; and, Maskeliunas, Damasevicius, Martisius, & Vasilievas, 2016). The Emotiv EPOC+ is an EEG headset with 16 wet electrode sensors (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4, M1, and M2) mapped to the scalp according to the standard 10-20 electrode layout (Figure 2). M1 (CMS) and M2 (DRL) act as reference points during the recordings. In order to create a strong connection between a participant's scalp and the electrode sensors on the headset, the Emotiv EPOC+ has gold-plated electrodes which are covered in felt pads that need to be wetted with saline solution. The wet felt pads act as a conduit between the scalp and the electrode sensors.

To record EEG data, the headset was connected wirelessly via Bluetooth to a Dell Inspiron 15 running Emotiv's proprietary data acquisition software, EmotivPRO, on Windows 10 Pro. EEG was recorded at a sampling rate of 128 Hz. In order to ensure a strong connection between the participant and the headset, a measure of contact quality native to the EmotivPRO software was referenced. Once an indication of 100% contact quality was ensured for all electrodes, the experiment began. Contact quality was continually monitored throughout the experiment. In order to limit head movement and reduce the impact of motion artifacts,

Figure 1. Emotiv EPOC+ headset.

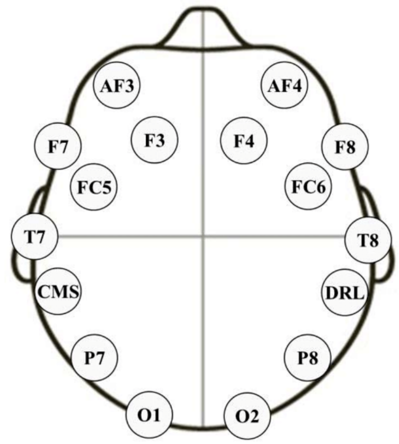


Figure 2. Layout of Emotiv EPOC+ electrodes.

participants positioned their head in a chin rest at a fixed height and distance from the screen (visual angle approximately 12.6º).

The main behavioral experiment was presented on a Dell XPS13 laptop running Ubuntu 16.04, which was connected to the data acquisition Windows laptop through a USB to Serial Port adapter cable. In order to analyze the EEG data stream at critical epochs, trigger codes with values that corresponded to different events in the experiment were sent from the Linux laptop via the USB to Serial Port cable. The trigger codes were received by a virtual COM port on the Windows laptop and embedded in the EmotivPRO EEG recording as event markers. These markers allowed for subsequent epoching of the data for trial-based feature analysis.

EEG Preprocessing. Electrophysiological data were preprocessed using Fieldtrip (Oostenveld, 2011). Scripts for data analysis were adapted from scripts used to generate results and figures in Rafidi et al. (2018) and inspired by scripts used in Rafidi (2012), Hirschstein (2018), and Dr. Joseph DeSouza's tutorial for preprocessing of Emotiv EPOC+ EEG data (DeSouza, 2014). Preprocessing steps were additionally guided by Dr. Steven J. Luck's suggestions for the order and nature of the processing of Event-Related Potential (ERP) EEG data (Luck, 2014).

Files were initially exported in European Data Format (.EDF format) from EmotivPRO. A program was written in MATLAB to prepare the atypical trigger code channel location in the .EDF file such that key events could be read into the Fieldtrip-specific data structures. Each .EDF file was initially treated as a single, continuous trial. This trial was then epoched into smaller, separate trials by searching in a trigger-channel for trigger codes, which co-occurred with stimulus onset times. When a trigger code was identified, a trial was defined as occurring 500 ms before the onset of the trigger to 2000 ms after.

A custom layout file was constructed to inform Fieldtrip of the Emotiv EPOC+ electrode locations. Electrodes were re-referenced to all channels. Channel frequencies for each trial were sent through a 0.1-30 Hz Butterworth bandpass filter (see Badcock, 2013, 2015 for filtering of Emotiv EEG data). A 60 Hz notch filter was applied in EmotivPRO before data exportation.

In order to preserve uniform trial lengths and not remove potentially critical components, a trial rejection approach was used in lieu of an artifact correction approach. Each individual trial was visually inspected for eye blinks, eye movements, and noise artifacts, and trials were rejected if these artifacts were identified. Z-scores were calculated for each sample-point in a given trial in order to highlight samples with particularly deviant frequencies. Z-scores greater than four standard deviations above the mean were flagged as potential artifacts. The Z-scores were used only as an instrument to aid in the trial rejection process — all trials were inspected manually in case artifacts slipped past the filter or the Z-scores picked up on seemingly regular oscillations. After going through the trial rejection process, across participants, an average of

17% of *Animal Family Classifier* trials were rejected, 13% of *Show and Tell Task* trials were rejected, 15% of *Swahili/English Task* trials were rejected, and 17% of *Semantic Category-Word Classifier* trials were rejected.

Classification Technique. EEG data from the classifier training paradigms (*Episodic Animal Family Classifier* and *Semantic Category-Word Classifier)* were used to train logistic regression models to classify high- and low-competition EEG voltages. A generalized logistic regression model determines the relationship between a set of input features (e.g. EEG channel voltages) and a categorical output variable (e.g. high- or low-competition). The classifier training paradigms included complete sets of labeled voltage features that correspond to binary competition output variables, which were used to train the models. An optimal time window for classification was previously established for the *Semantic Category-Word Classifier* (Rafidi et al., 2018) by looking at average voltage activity across different post-stimulus time points. An analogous search process was conducted for the *Episodic Animal Family Classifier* by averaging voltage activity across 50 ms time windows. The classifier training process was conducted at each time window to identify the time window with the best classification accuracy.

For this experiment, the training dataset was generated from the classifier training paradigms and, for each time window, took the form of a matrix (samples by channel features) with a binary indicator identifying whether each sample was a high- or low-competition retrieval trial. For a given participant, a classifier was trained by feeding this matrix into a logistic regression model and receiving feature weights as an output. These weights represent the relative variance accounted for in the model by each channel. In order to determine the accuracy of the classifier, these data were separated into five cross-validation folds. The classifier was trained on four folds and tested on the fifth, creating a time series of classification accuracies across all time windows for each participant. The time series was then averaged across participants to find the best overall classification window. Significance and reliability at each time point were measured using the same method outlined in Rafidi et al (2018).

Once the feature weights were established for a given participant, the logistic regression model could be used to make predictions about which category a trial from a new set of data belongs. Applying the classifier to data from a classifier testing procedure (either the *Swahili/English Task* or the *Show and Tell Task*) generates predictions about which category a given trial belongs to, based on the features that occur in the dataset. The classifier learns from the training set what high- and low-competition voltages should look like. When given a testing dataset, the classifier searches in the pre-specified time windows to find whether the voltages look more like high-competition or low-competition voltages. It then makes a prediction, based on the relative variance that each channel feature accounts for, about whether there is high- or low-competition in the participant's brain during a given trial.

Stimuli

Stimuli used in the *Semantic Category-Word Classifier* and *Swahili/English Task* were the same materials used in Rafidi et al. (2018). Stimulus presentation orders were generated and randomized before the beginning of the experiment for each participant.

Images for the animals used in the *Episodic Animal Family Classifier* training phase were drawn from the stimulus set created by Hulbert and Norman (2015), and additional photos were sourced from the Animals with Attributes2 dataset (Xian, Schiele, & Akata, 2017). The final animal images dataset included the following ten animal groups: Antelope, Bear, Elephant, Fox,

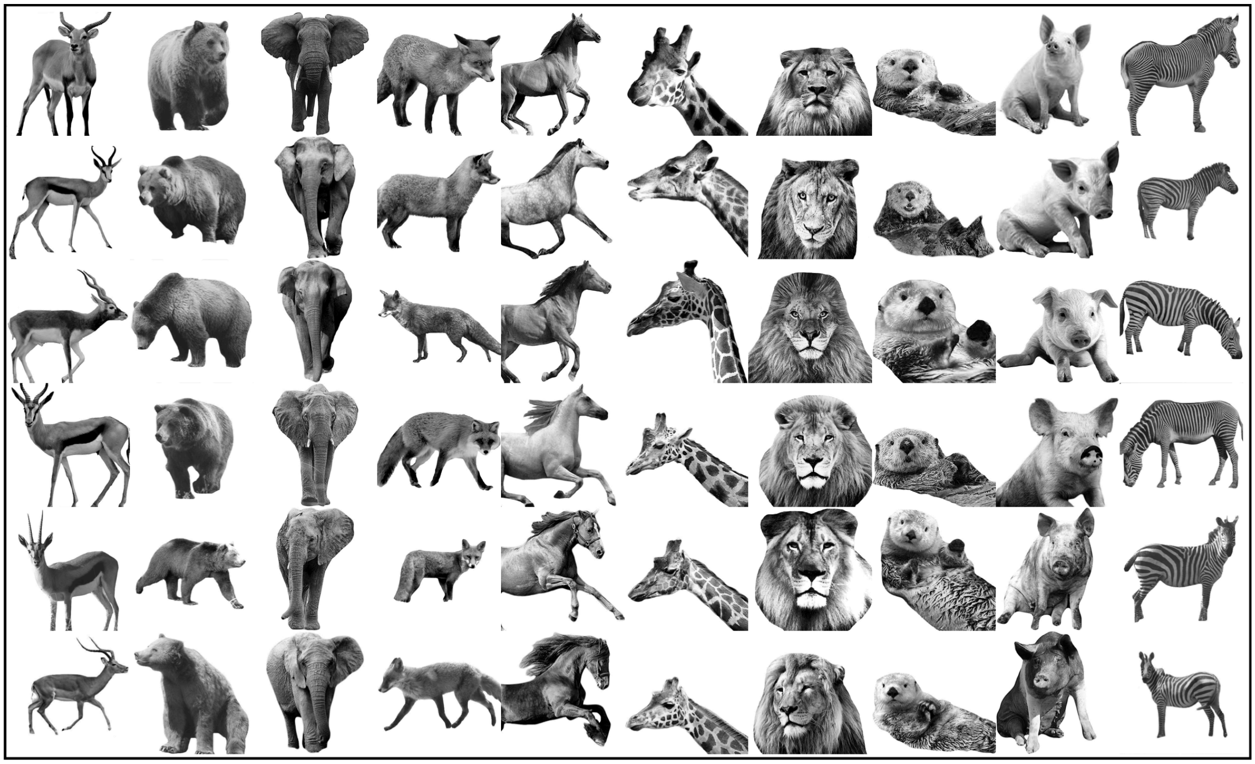


Figure 3. Image stimuli used in the *Episodic Animal Family Classifier*.

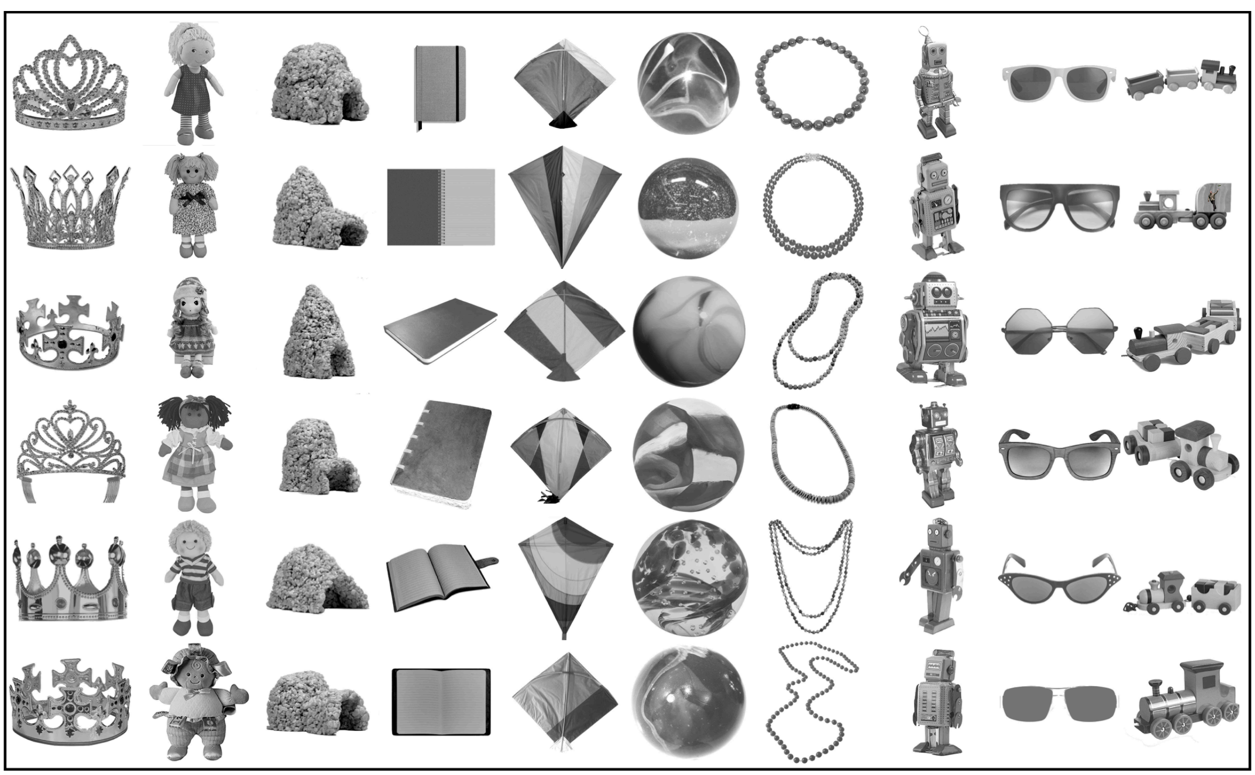


Figure 4. Image stimuli used in the *Show and Tell Task*.

Giraffe, Horse, Lion, Otter, Pig, and Zebra (Figure 3). Images for the preschoolers' items used in the *Show and Tell Task* were sourced from the Internet using Google Images. The final item dataset included the following ten item groups: Crown, Doll, Igloo, Journal, Kite, Marble, Necklace, Robot, Sunglasses, and Train (Figure 4). The item images dataset was designed to mimic the factors that were considered while generating the animal images dataset; within item groups, items were visually distinct, but difficult to distinguish to the untrained eye.

Images for both image datasets were converted into 500 by 500 pixel .png files using Apple Preview, such that the backgrounds were removed for all images, and the animal or item in the foreground remained. Images were then converted into .jpg files and preprocessed in MATLAB using the Shine Toolbox (Willenbockel et al, 2010) to match the luminance histograms across all images in the dataset. The resulting datasets were comprised of grayscale images.

Names for the animals in the *Episodic Animal Family Classifier* and the preschoolers in the *Show and Tell Task* were sourced from the United States Social Security Administration (SSA) database. The SSA records contain a file for each year of birth between 1879 and 2017, listing data for all people born during each year who received a Social Security Number: names and the frequency of each name, as well as the sex associated with each name. R, an open-source software environment for statistical computing (R, 2013), was used to filter the names from the SSA records to find six-letter names that were relatively infrequent (between 5 and 50 occurrences) in the year 1997, a year in which a large portion of the participant pool at Bard College would have been born. Infrequently used names were chosen such that participants would not have strong associations with most of the names. Any associations that participants

Antelope	Bear	Crown	Doll	Elephant	Fox	Giraffe	Horse	Igloo	Journal
Aldair	Booker	Cormac	Dorain	Elbert	Felton	Gerrit	Hisham	Irvine	Jabril
Ashten	Briant	Curran	Dustan	Eshawn	Furkan	Gunter	Hunner	Isaac	Jerell
Andrei	Bailee	Caston	Decker	Easten	Frandy	Gilmar	Hykeem	Imraan	Juwaan
Arline	Bethel	Chayla	Daylin	Ellyse	Finley	Grisel	Hailee	Ingris	Jissel
Austyn	Bintou	Cyndel	Dylann	Erynne	Foster	Goldie	Hollin	Itzell	Jhayla
Azlynn	Blayke	Claira	Diedre	Evette	Farren	Gaelyn	Hettie	Ivorie	Jolena
Kite	**Lion**	**Marble**	**Necklace**	**Otter**	**Pig**	**Robot**	**Shell**	**Train**	**Zebra**
Kyland	Landin	Mychal	Noland	Oakley	Parish	Randon	Shamus	Trevis	Zuhayr
Koltin	Lucien	Murray	Newell	Osmond	Puneet	Rustin	Stevin	Thayne	Zoltan
Khayla	Lorenz	Monroe	Nuchem	Othman	Phelan	Rizwan	Salman	Tyrome	Zander
Karsyn	Lilith	Merrit	Nycole	Odette	Prisca	Roslyn	Sidnee	Torrey	Zissel
Kirsti	Lynsie	Mirsha	Nadyne	Oonagh	Porcha	Rhiley	Skylyn	Taylee	Zeinab
Kellan	Leilah	Maddux	Nishat	Orchid	Perrin	Reegan	Sondra	Tenzin	Zhanee

Figure 5. Stimuli for *Episodic Animal Family Classifier* and *Show and Tell Task*. Two or six names were randomly chosen from each letter group across the two tasks.

might have with these infrequently used names would likely pan out as noise. Sixty names in total were taken from the database (Figure 5). Ten groups were created from these sixty names. Within each group, all names began with the same letter and had unique second letters.

For each participant, half of the animal images and animal names were assigned to be two-member groups and half were assigned to be six-member groups. In the six-member groups, all six images and all six names were randomly assigned to one another. In two-member groups, two names and two images were randomly pulled from the stimulus pool and paired together. Separately, item images and item names were also divided into two-member and six-member groups using the same method of randomization.

Procedure

The experiment followed a within-subjects design and took place over the course of three sessions, separated each time by an unconstrained delay of one week (the third session took place exactly seven days after Session Two for all but one participant, for whom Session Three took place nine days after). Across the three sessions, every participant went through two classifier training procedures and two classifier testing procedures (Figure 6).

Session One. All participants began the first session of the experiment by signing an informed consent form and completing a basic demographics form. Participants were then outfitted with the Emotiv EPOC+ and wore the headset throughout the duration of the experiment. The headset was taken off between tasks or between blocks within a task only if electrode contact quality was shown to be poor and inconsistent, or if the participant noted that the headset was painful. Whenever the headset was taken off, saline solution was reapplied to the sensors to create a stronger contact quality and ensure a more comfortable experience for the

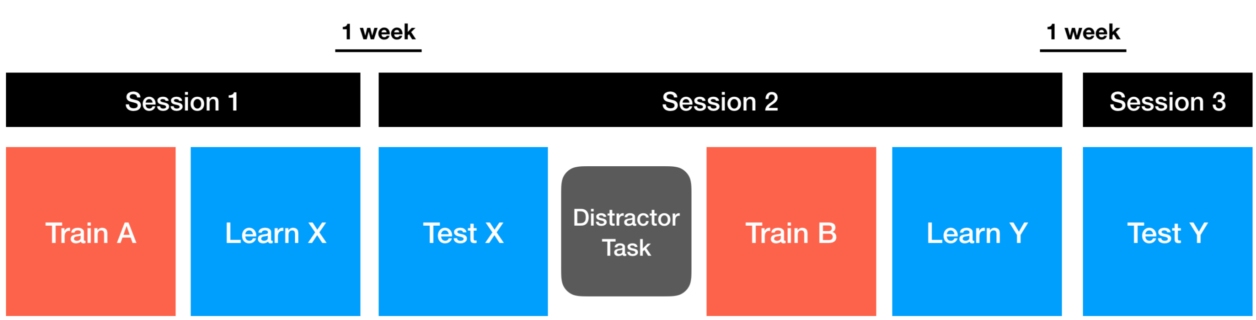


Figure 6. General overview of experimental design and task counterbalancing. Red blocks could be either classifier training procedure (either *Episodic Animal Family Classifier* or *Semantic Category-Word Classifier*), and were counterbalanced across A and B positions. Pairs of blue blocks refer to classifier testing procedures (either *Swahili/English Task* or *Show and Tell Task*), which were counterbalanced across X and Y positions. Task order was fully counterbalanced across all eight participants. See Figure 7 for counterbalanced task orders for each individual participant.

participant. While performing tasks during the experiment, participants positioned their head in a chin rest attached to the table in front of them. After completing a classifier training procedure and a classifier testing procedure, participants were given a questionnaire asking what percentage of the learned material, on a sliding scale of 0 to 100, they expected to remember at various points in the future (Figure 7). Participants were thanked, asked to return to the same testing room one week later, and dismissed.

Session Two. In the second session, participants were welcomed back to the same room as was used in their first session and were tested on the material that they learned during the classifier testing procedure in the first session. The testing phase was conducted using Google Forms for both classifier paradigms. After completing the testing phase, participants were given a two-minute distractor task (Appendix 3) in which they were asked to do their best to solve a 9x9 Sudoku puzzle. Participants were then outfitted once again with the Emotiv EPOC+, asked

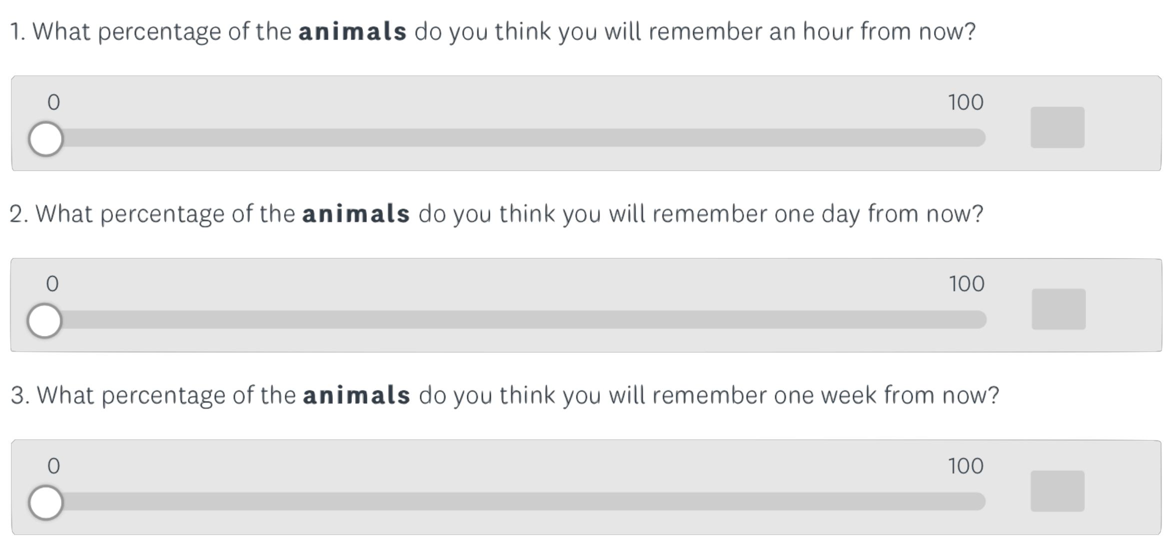


Figure 7. Post-task questions delivered at the end of Session One and Session Two. Participants responded to these three recall-delay questions for all four tasks by the end of the experiment: *Episodic Animal Family Classifier* (shown above), *Show and Tell Task*, *Semantic Category-Word Classifier*, and *Swahili/English Task*.

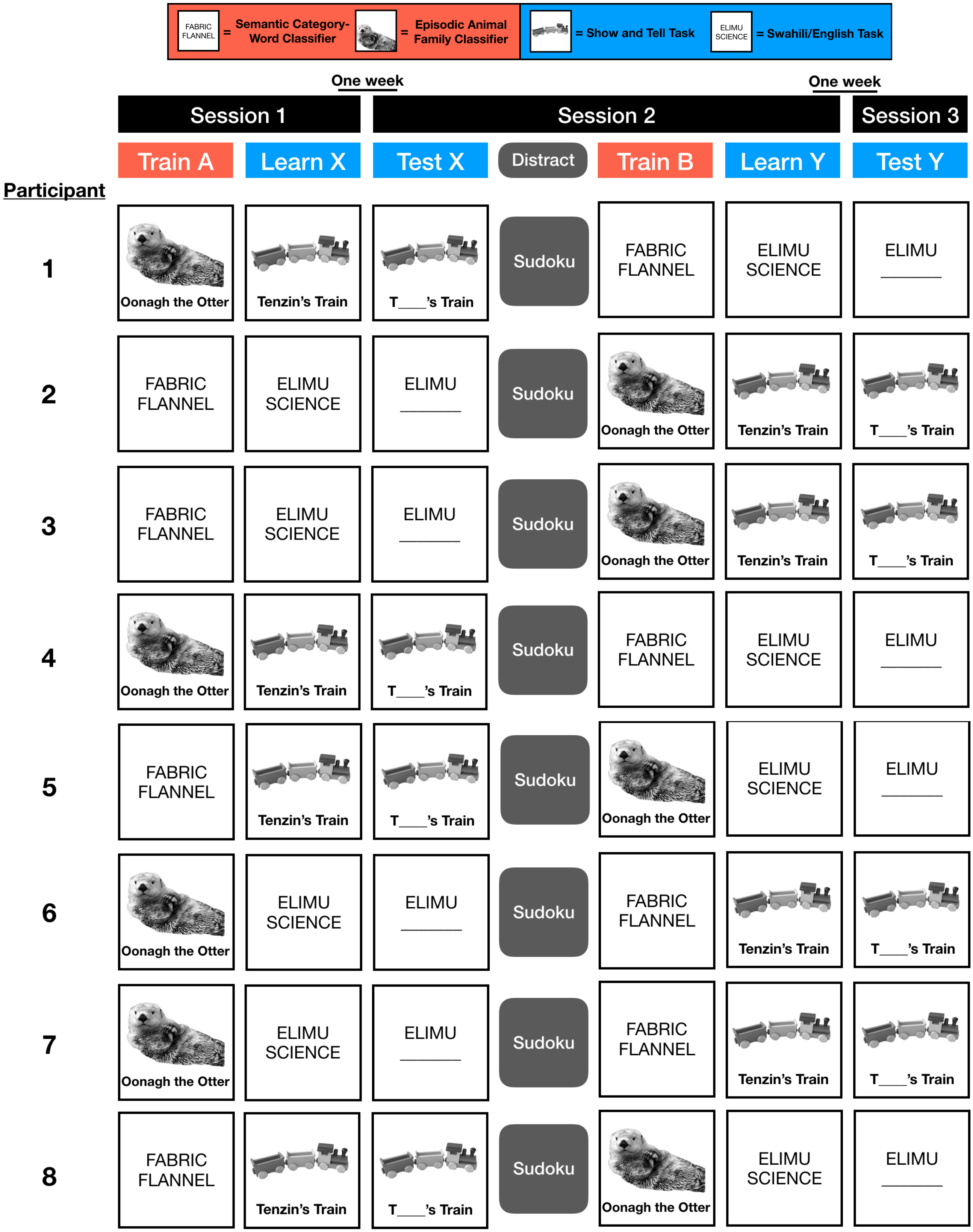


Figure 8. Outline of procedure counterbalancing for all participants. Participants one through four were run before spring intersession, participants five through eight were run after. See Figure 6 for general overview of procedure counterbalancing.

to rest their head on the chin rest, and began the classifier training procedure they did not complete during the first session. Participants then completed the second classifier testing procedure while still connected to the Emotiv EPOC+. Finally, the headset was removed and participants completed a questionnaire asking how much learned material they expected to remember in the future (Figure 7). Participants were thanked, asked to return to the same room one week later, and dismissed.

Session Three. In the third and final session, participants were welcomed back to the same room and were given the final testing phase, in the format of a Google Form, for the material that they had studied during the classifier testing procedure in the Session Two. Participants did not wear the Emotiv EPOC+ during Session Three as no EEG data was recorded. Immediately after completing the final cued recall test, participants completed a Post-Experiment Questionnaire. Participants were then debriefed, thanked and paid for their time and participation, and dismissed.

Classifier Training Procedures

Participants completed training procedures for two unique classifiers: the *Episodic Animal Family Classifier* and the *Semantic Category-Word Classifier*. Classifier training procedures established and manipulated different types of competition (Figure 9).

Semantic Category-Word Classifier. The *Semantic Category-Word Classifier* paradigm was a replication of the method used in Rafidi et al. (2018). This classifier was used as a comparison to assess the accuracy and strength of the *Episodic Animal Family Classifier*.

The classifier was trained using 60 category-exemplar word pairs (Appendix 2). Each category was associated with eight exemplars, each of which began with a unique letter, within

Features of Competition Manipulated in Task

Semantic Category-Word Classifier	Episodic Animal Family Classifier
Features of semantic category-word relationship	Number of animals in a species enclosure
	Time allotted for recall during retrieval-practice blocks
	Visual features
	Orthographic features

Figure 9. Outline showing the facets of competition that each classifier is designed to identify.

category. Participants completed one block of study trials, during which they were consecutively presented with a randomized order of category-exemplar pairs (e.g. TREE-PINE). Each pair was presented for two seconds in the center of the screen and was followed by a fixation cross presented for one second.

Following the study block, participants completed four blocks of retrieval-practice for the studied category-exemplar pairs (Figure 10). During retrieval-practice trials, participants were told to remain still until they knew the answer, such that the EEG data would not be corrupted by movement artifacts. In what was referred to as a low-competition trial, participants were cued with the first letter of the category and an exemplar, and they were tasked with retrieving the full name of the category (T-PINE). In what was designated as a high-competition trial, participants were cued with the full name of the category and the first letter of an exemplar, and they were tasked with retrieving the full name of the exemplar (TREE-P). During the retrieval-practice blocks, participants were told to think of the answer in their head and press the spacebar once they had thought of the answer. As participants did not provide any form of recorded answer, there is no way to officially verify whether their responses were correct. However, given the pre-existing semantic associations that were trained during the study phase, it was assumed that

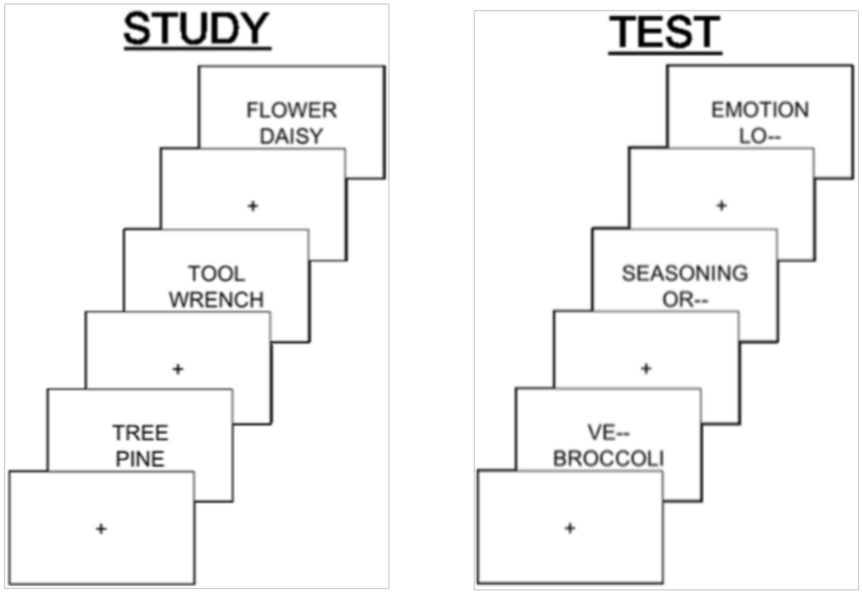


Figure 10. Overview of study and test blocks for the *Semantic Category-Word Classifier.* Participants completed one block of study trials followed by four blocks of test trials. *Note.* Adapted from Rafidi et al. (2018).

participants would be fairly accurate and experience more competition when attempting to recall the exemplars given the categories. Trials continued automatically after three seconds whether or not a participant pressed the spacebar.

Episodic Animal Family Classifier. The *Episodic Animal Family Classifier* was trained using a novel experimental classifier training task which aimed to identify distinct states of high- and low-competition. By establishing these levels of competition a priori in the set of training materials, the classifier captures competition signals at the time of encoding.

In the *Episodic Animal Family Classifier* training procedure, participants learned a set of associations formed in the laboratory. In this way, none of the material was reliant upon any pre-existing semantic associations. Participants were introduced to a scenario in which they were working at a zoo. Their task was to learn the names of all of the animals to the point where they would be able to remember the name of an individual animal when presented with its photo later on. Animals were organized by species into different enclosures and all of their names began with the same letter as the first letter of species (e.g., all of the Zebras were in one enclosure and all of their names began with the letter, Z).

Competition was manipulated across many dimensions. Within a species, the visual features of the animal images were very similar, but slightly different, making the animals difficult to distinguish and remember to the untrained eye. Additionally, all of the animal names in an enclosure began with the same letter, were all six letters long, and had the same frequency of use, adding several levels of orthographic competition which required intensive training to learn the differences. Competition was additionally operationalized as the number of similar animals that were in a species' enclosure. Low-competition was tagged for enclosures with only two animals (e.g. the Lion enclosure: Landin and Lorenz the Lions). High-competition was tagged for enclosures with six animals (e.g. the Otter enclosure: Oakley, Osmond, Othman, Odette, Oonagh, and Orchid the Otters). In this way, the level of competition was created a priori within the set of material in the experimental design and should be relatively consistent across participants at the time of encoding, which occurred during the experiment. Finally, during the retrieval-practice blocks, participants were allotted a shorter amount of time to recall the low-competition animals than they were given to recall the high-competition animals. The shorter amount of time required participants to have learned low-competition animals to a greater extent, such that their recall reaction times were shorter. By considering so many dimensions of competition, the difference between high- and low-competition would be maximized, making

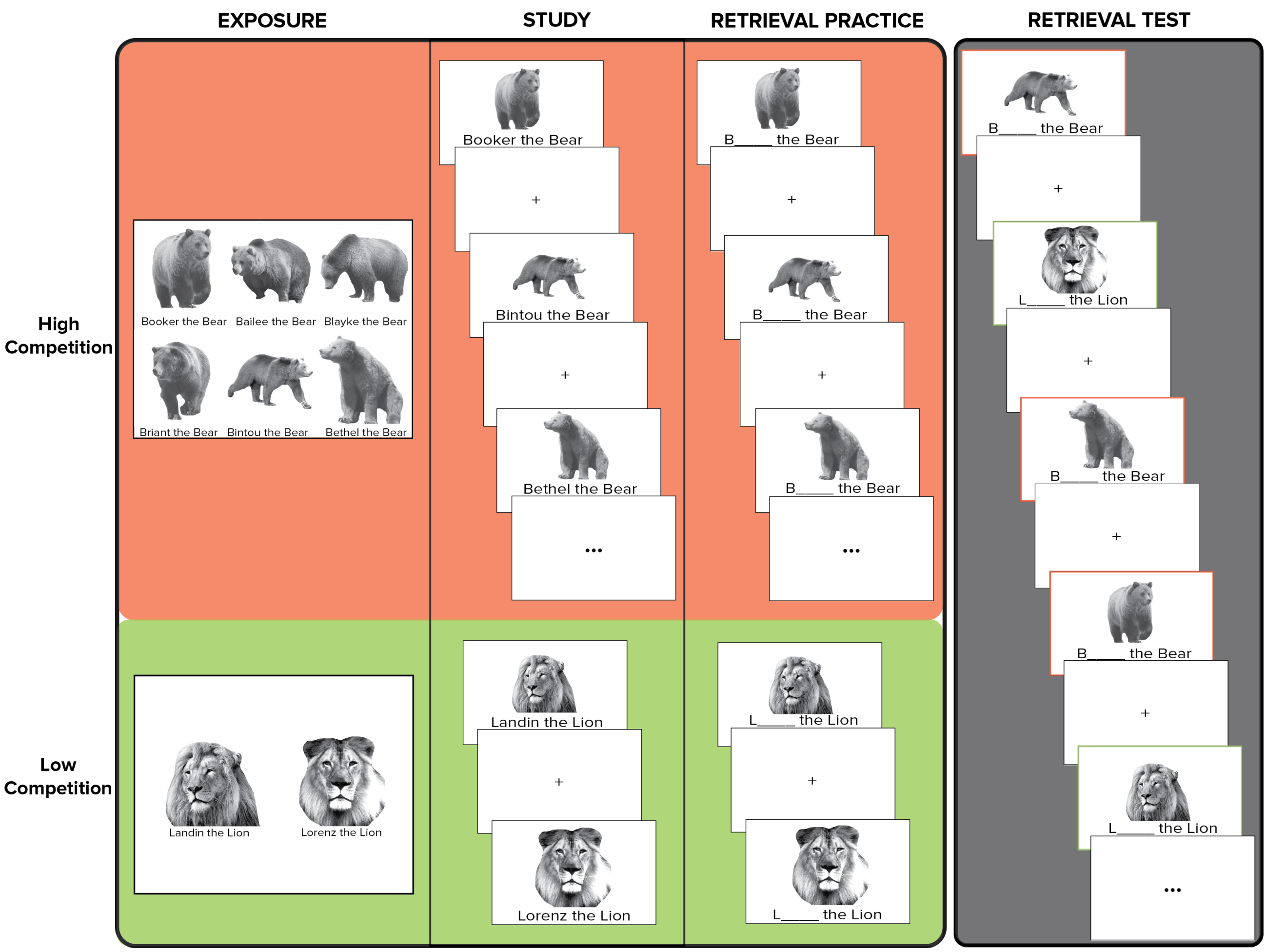


Figure 11. General overview of the *Episodic Animal Family Classifier* training procedure. Shown are the exposure, study, and retrieval-practice blocks for high- and low-competition animals. Participants completed all three blocks for ten animals. Also shown is the final cued-recall retrieval test that participants completed for all animals grouped together.

classification more clear, while also representing a more naturalistic learning environment that does not manipulate variables on a single dimension.

Participants completed three different blocks for each animal enclosure in the zoo, one animal at a time: exposure, study, and retrieval-practice (Figure 11). In an exposure block, participants were introduced to all of the animals in a given species at the same time. The images of all of the animals were shown and their names were presented below the images. In a study block, participants were shown, one after the other, a complete pairing of the image of an animal and the animal's name below the image. After all of the animals in an enclosure were studied, participants moved on to the retrieval-practice block.

In a retrieval-practice block, participants were presented with an image of an animal and tasked with trying to recall the name of the given animal. As each image appeared on the screen, participants were given two seconds to think about the name of the animal. They were told that it it was important to remain still and not blink during the two second period, such that movement artifacts would not contaminate the EEG acquisition. After the two seconds had elapsed, a prompt with the first letter of the name appeared on the screen and participants had to type the full name of the animal using the laptop keyboard. They could start typing as soon as the text appeared. If the participant was unsure about the answer and did not make a response after one second had passed, one more letter was revealed every second until a pre-specified proportion of the word was revealed. For high-competition trials, a name was considered to be unlearned if the participant did not type the correct answer before a total of three letters were revealed (i.e., the participant failed the trial if they could not produce the answer when given half of the letters in the animal's name). For low-competition trials, the name was considered to be unlearned if the participant did not type the correct answer before two letters were revealed. To stop the reveal of letters and make a guess, participants pressed the spacebar. If the participant did not type the name before the trial-specific number of letters were revealed, the trial was failed and it was considered to have not yet been learned. If the trial was failed, the whole word was then uncovered and the participant was given a chance to restudy the name/image pair for five seconds. Failed trials were put back into a list of unlearned trials and retrieval-practice was repeated until they were accurately recalled using the parameters outlined in this drop-off procedure.

The exposure, study, and retrieval-practice blocks were designed for the participants to learn the stimulus set. The purpose of these blocks was to build up a stable level of competition difference across the high- and low-competition animal groups, such that by the time that participants reached the final retrieval test block, competition levels were established and distinct for the different animal groups.

After the exposure, study, and retrieval-practice blocks were completed for all of the animals, participants moved on to a final retrieval test block. A prompt appeared on the screen explaining that all of the animals in the zoo escaped from their enclosures and they needed to be identified. All of the animal images were then presented once in a randomized order. For each animal, there was a two second EEG acquisition period during which participants could not type their response. Data from the EEG acquisition period during the final retrieval test block was used as the training data for the classifier. After the two second acquisition period elapsed, participants had to recall the name of each animal with only the first-letter prompt, typing their response using the laptop keyboard in front of them.

Classifier Testing Procedures

Data collected during the classifier testing procedures were used to test the effectiveness of the *Episodic Animal Family Classifier* and the *Semantic Category-Word Classifier*.

Swahili/English Task. The *Swahili/English Task* was the original task in Rafidi (2018) upon which the effectiveness of the *Category-Item Classifier* was tested. Data from this task comprised a testing dataset that both classifiers were tested on. If the *Episodic Animal Family*

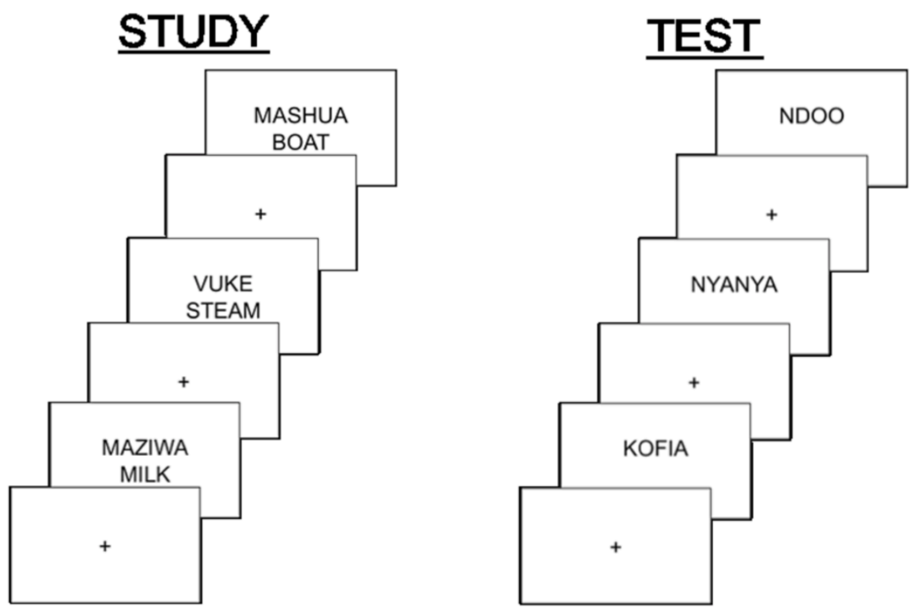


Figure 12. Overview of study and test blocks for the *Swahili/English Task*. Participants went through four blocks of study trials and four blocks of test trials. Blocks were interleaved. *Note.* adapted from Rafidi et al. (2018).

Classifier is better than the *Semantic Category-Word Classifier* at predicting performance on the *Swahili/English Task*, this would suggest that the *Episodic Animal Family* C*lassifier* is particularly attuned to picking up on competition, given the results from Rafidi et al. (2018) which demonstrated significant classification accuracy for the *Semantic Category-Word Classifier*.

In the *Swahili/English Task*, participants completed eight blocks of interleaved study and retrieval-practice trials for Swahili/English word pairs (Figure 12). In a given study block, participants would watch on a computer screen as 60 Swahili/English word pairs appeared in a randomized order, one after another. A pair appeared on the screen for two seconds and was followed by a one second fixation cross. In a given retrieval-practice block, participants were presented with a Swahili word and asked to recall the English translation by typing their response on a keyboard. If they did not know the word, they could press “enter” to skip the

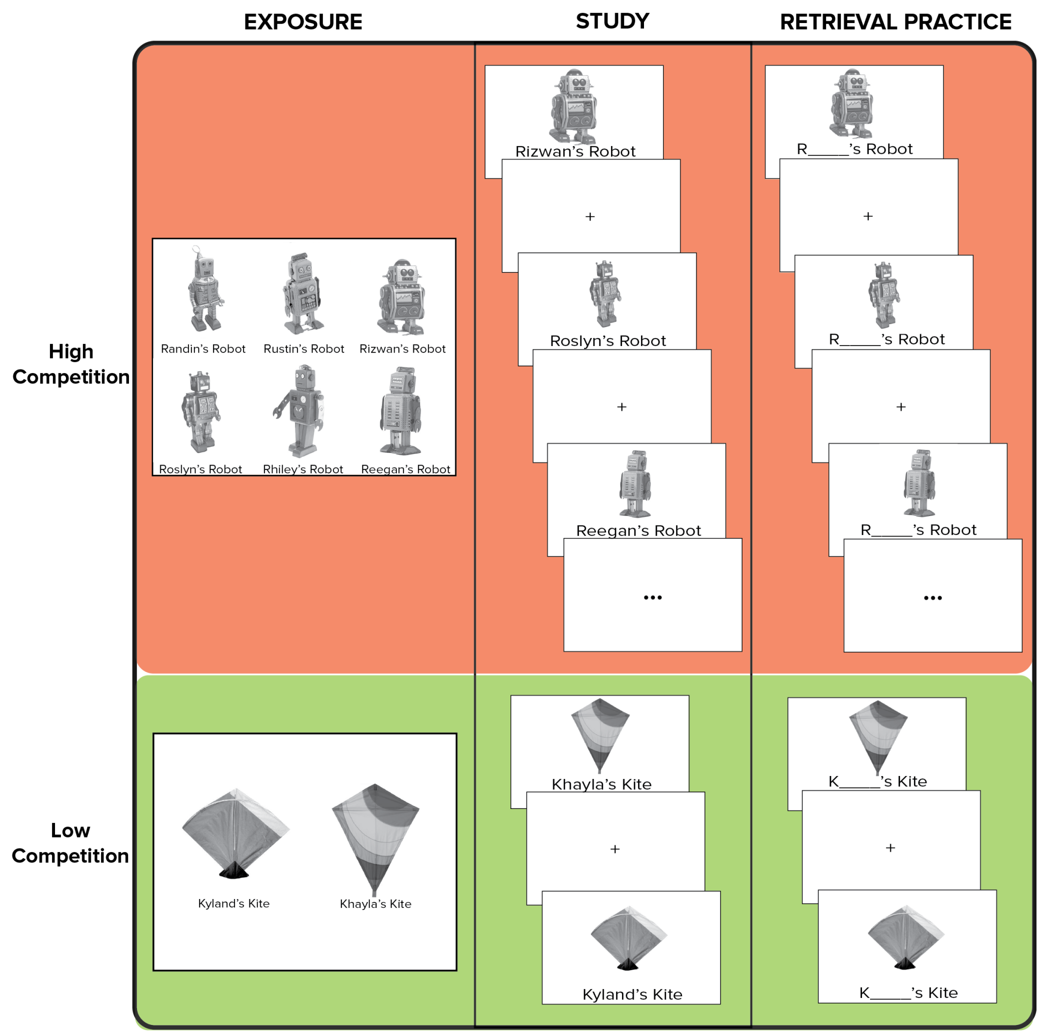


Figure 13. General overview of the *Show and Tell Task*. Shown are the exposure, study, and retrieval-practice blocks for high- and low-competition items. Participants completed all three blocks for ten items.

prompt. Participants were told to remain still for two seconds after the prompt appeared on the screen, as the recorded EEG data from the retrieval blocks of the *Swahili/English Task* would be used as a testing dataset for each classifier.

Exactly one week later (with the exception of one participant who was tested nine days after initial learning), participants returned to the lab and were tested on their memory for the learned Swahili/English word pairs. No EEG data was collected during the final recall phase and thus the participants were not outfitted with the Emotiv EPOC+. Participants were given 15 minutes to complete a final cued-recall test on Google Forms that prompted them with Swahili words and required them to fill in the English translations. They could move back and forth between questions freely, and the experimenter waited until the full 15 minutes had elapsed to continue with the experiment.

Show and Tell Task. Data from the *Show and Tell Task* were used as a testing dataset for both classifiers. Because the *Show and Tell Task* and the *Episodic Animal Family Classifier* training procedure have parallel paradigms and deal with similarly structured study material, the *Episodic Animal Family Classifier* will likely be able to effectively predict performance on this task.

In the *Show and Tell Task*, participants were told to imagine that they were a preschool teacher on Show and Tell Day (Figure 13). On Show and Tell Day, students were invited to each share a specific item that begins with the same letter as their name. As their teacher, participants needed to remember which item was chosen by which student. Similar to the *Episodic Animal Family Classifier*, participants completed exposure, study, and retrieval-practice blocks in order to learn student's names and their items. For each group, they would first see all of the student's names and their items presented together as a group. They then saw each student's name and their item individually. Finally, participants completed the retrieval-practice block in which they were shown an image and had to guess the name of the student to which it belonged.

As each image appeared on the screen, participants were given two seconds to think about the name of the student associated with the item. They were instructed to remain still during this two seconds, as recorded EEG data would be used to later test the predictive capability of the classifiers. After the two seconds had elapsed, a prompt with the first letter of the student's name appeared on the screen and participants had to type the full name using the laptop keyboard. If a participant was unsure about the name of the student, one more letter was revealed every second until a pre-specified proportion of the word was revealed. As with the the *Episodic Animal Family Classifier*, for high-competition trials, a name was considered to be unlearned if the participant did not type the correct answer before a total of three letters were revealed (i.e., the participant failed the trial if they could not produce the answer when given half of the letters in the student's name). For low-competition trials, the name was considered to be unlearned if the participant did not type the correct answer before two letters were revealed. Participants were instructed to press the spacebar to stop the reveal of letters to begin typing if they knew the response. If they did not type the name before the pre-defined amount of letters were revealed, the whole word was uncovered and the participant was given a chance to restudy the name/image pair for five seconds. Failed trials were placed back into a stack of unlearned trials and re-tested until the participant was able to accurately recall the name a single time under the described conditions and pressures of the letter reveal.

One week later, participants returned to the lab and were tested on the names of the preschool students. No EEG data was collected during the final recall phase and thus the participants did not wear the Emotiv headset. Participants were given 15 minutes to complete a Google Form that presented an image of an item and required the participant to type the name of the student associated with the item. They could move back and forth between questions freely, and the experimenter waited until the full 15 minutes had elapsed to continue with the experiment.

Results

The complete set of intended analyses were not able to be sufficiently run. Due to an error realized only several days before the Senior Project submission date, EEG data from the *Swahili/English Task* and the *Semantic Category-Word Classifier* were corrupted. Efforts are being made to resolve this issue in order to effectively run the full, multi-classification analysis suite. Once the analysis pipeline has been updated and the data issues resolved, analysis scripts will be committed to the previously mentioned GitHub repository. As a result, the effectiveness of the novel *Episodic Animal Family Classifier* can not be tested on the *Swahili/English Task* data, nor can it be compared to the intended replication of the Rafidi et al. (2018) *Semantic Category-Word Classifier*. Additionally, predictions about the fate of specific items in the *Show and Tell Task* are unable to be made. Nevertheless, results from behavioral data for all tasks and electrophysiological results from the *Episodic Animal Family Classifier* training procedure and the *Show and Tell Task* can be discussed in relation to the established literature.

Data for all eight participants were included in the final analyses. A bar for exclusion was set, in line with previous standards set by Luck (2014), such that a participant would be excluded from final analyses if more than 25% of the total trials (aggregated across tasks) were rejected during EEG preprocessing. No participants were excluded under these constraints or for any other reason. Welch's two-sample t-test for unequal variances revealed that there was no difference in recall for participants run before or after the spring intersession for either Classifier Testing Procedure (*Show and Tell Task*, $p = 0.49$; *Swahili/English Task*, $p = 0.18$), suggesting that counterbalancing the order of tasks did not significantly affect final recall. As such, analyses were collapsed across task counterbalancing conditions.

Behavioral Competition Check. For the *Episodic Animal Family Classifier*, participants completed retrieval-practice trials for each animal at least once, for a total of at least 40 retrieval-practice trials across all animals. The amount of additional retrieval-practice trials that participants completed, as a result of failing a trial, were compared between high- and low-competition animal trials. A two-tailed, paired-samples t-test revealed a significant difference between the mean amount of additional high-competition trials (M = 9.875) and low-competition trials (M = 1.5) that participants had to go through during the retrieval-practice phase, as a result of not correctly recalling trials under the defined constraints ($p < 0.001$), suggesting that high-competition trials took longer to learn.

A paired-samples t-test revealed a significant difference between the amount of high-competition (M = 5.875) and low-competition (M = 3.375) trials that participants correctly answered in the final retrieval test phase during the *Episodic Animal Family Classifier* training procedure ($p < 0.05$). Considering that there were a greater number of high-competition animals (30) compared to low-competition animals (10), a second paired-samples t-test was run to compare the relative percent correctness for high-competition (M = 0.196) and low-competition animals (M = 0.338), $p = 0.076$. Together, these two statistical tests suggest that while the raw values from the retrieval test showed greater recall for high-competition animals, normalizing the scales to compare relative correctness revealed a marginally significant trend towards a greater percentage of low-competition animals being recalled compared to high-competition animals. Future work will consider methodologies to introduce equivalent amounts of high- and low-competition animals so that raw recall can be directly compared.

Judgments of Learning. Participants' judgments of their own learning (JOLs) were measured at the end of Session One and Session Two when asked what percentage of the study material they believed they would remember after several time delays. Figure 14 shows judgments of learning at each time delay for all tasks for all participants. For all tasks, trend lines

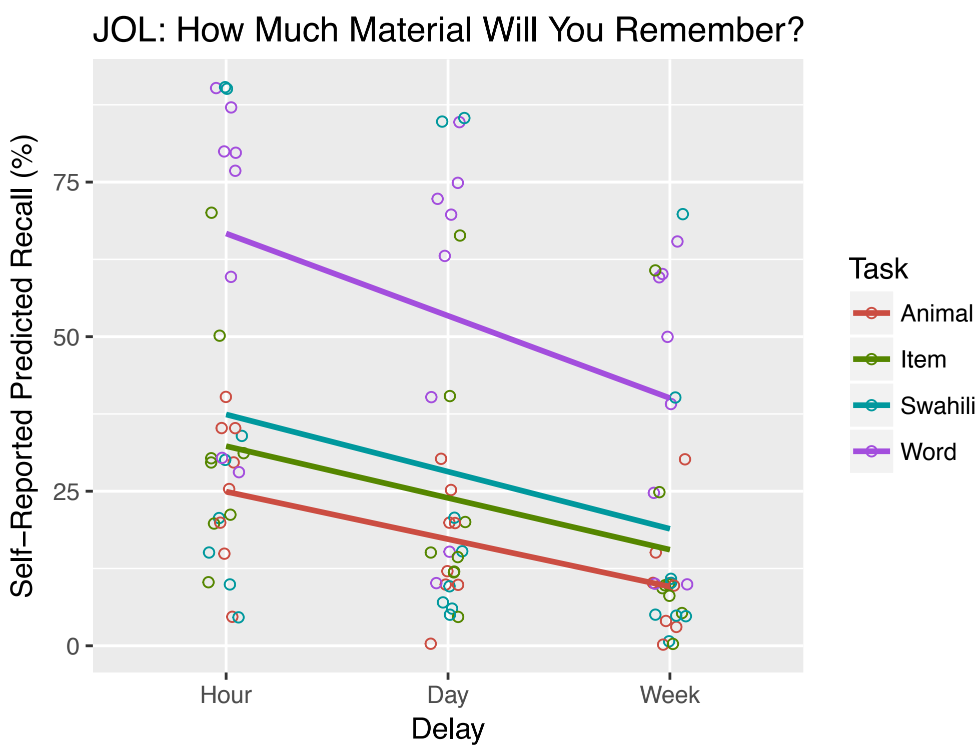


Figure 14. Judgments of learning for all participants for all tasks. Each point represents an individual participant's response for the given task and delay. Points are jittered on the horizontal axis to increase visibility of data. See Figure 7 for overview of questions asked.

display that participants believed their memory for learned material would decline as time increased. Two-tailed, paired-samples t-tests revealed a significant difference between JOLs for delays of one hour and delays of one week for the *Episodic Animal Family Classifier* training procedure ($p < 0.05$) and the *Semantic Category-Word Classifier* training procedure ($p < 0.05$). Additionally, collapsing across time delays revealed that participants rated they would recall a greater percentage of the material in the *Semantic Category-Word Classifier* training procedure than they would recall for the *Episodic Animal Family Classifier* training procedure ($p < 0.000001$), the *Show and Tell Task* ($p < 0.0001$), and the *Swahili/English Task* ($p < 0.01$).

Long-Term Recall for Classifier Testing Procedures. One week after initial learning, recall for the *Show and Tell Task* (M = 1) was near floor performance (Figure 15). One week after initial learning, final recall for the *Swahili/English Task* (M = 16.375) was greater than floor

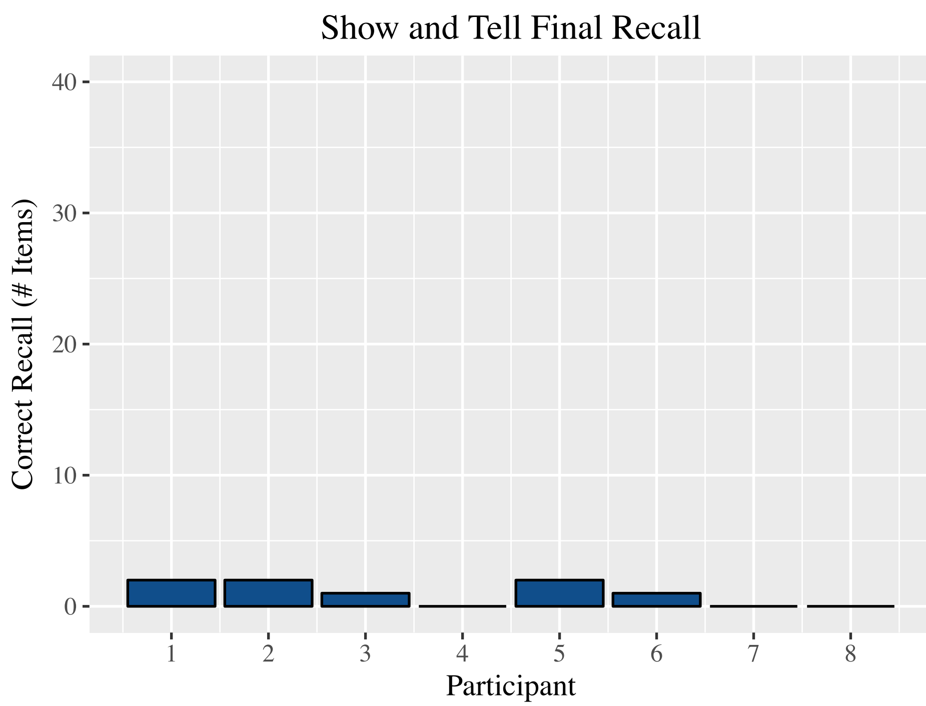


Figure 15. Final raw long-term recall for *Show and Tell Task* for each participant.

performance (Figure 16). To compare performance on the *Show and Tell Task* and the *Swahili/English Task*, the final cued-recall test was normalized by comparing the percentage of correct trials recalled for each task (Figure 17). A two-tailed, paired-samples t-test revealed that participants recalled a significantly greater percentage of *Swahili/English* items than *Show and Tell* items ($p < 0.01$).

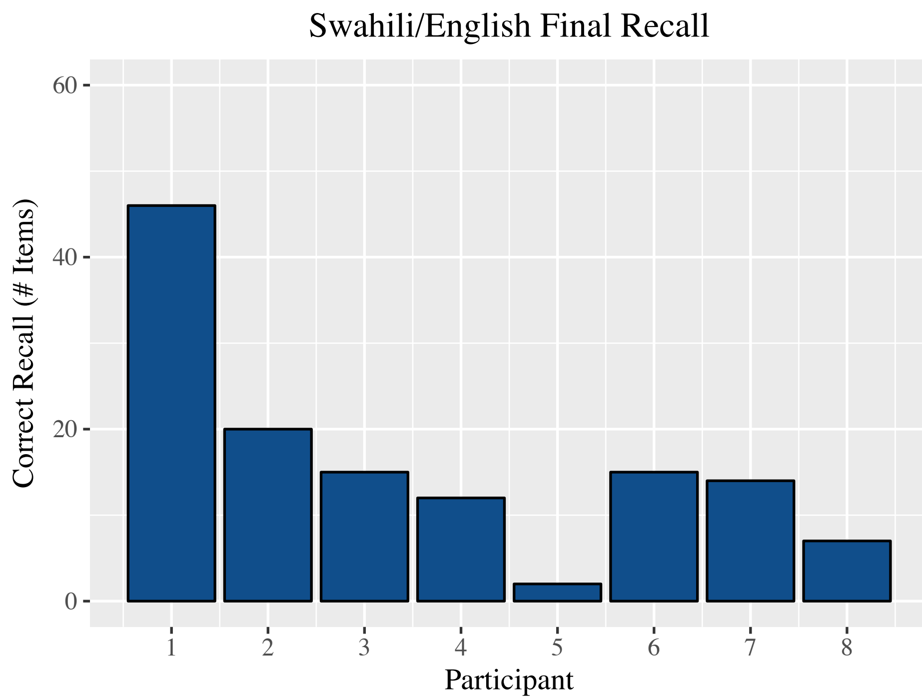


Figure 16. Final raw long-term recall for *Swahili/English Task* for each participant.

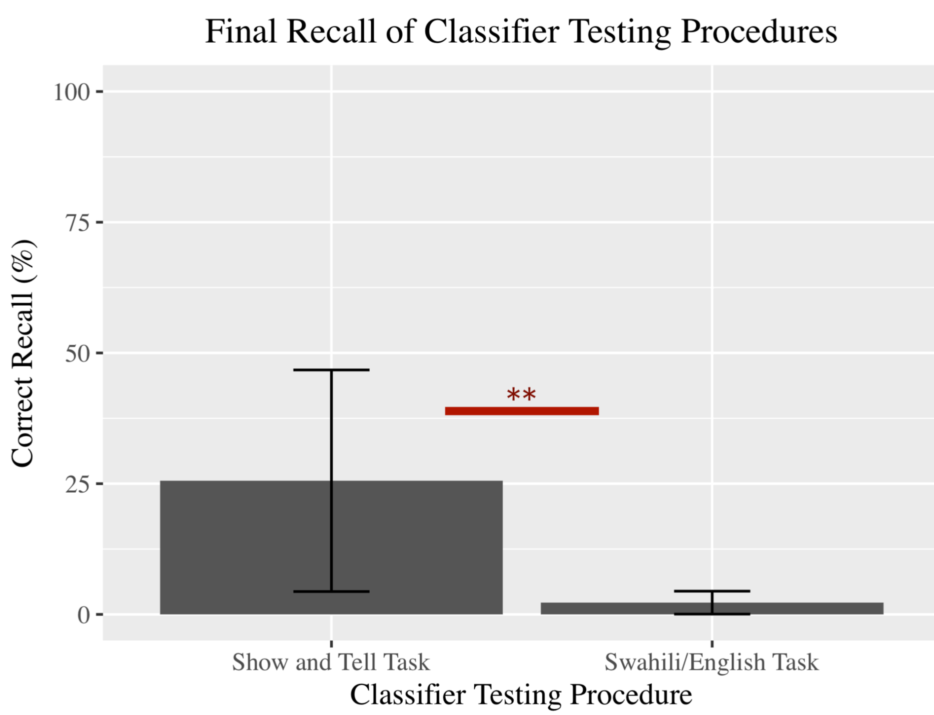


Figure 17. Mean percentage correct recall across participants for both classifier testing procedures during final recall one week after initial learning. Significant at $p < 0.001$. Error bars = SD.

Competition Classifier Accuracy. Competition decoding accuracy details the percentage of time that the classifier will accurately predict whether there is high- or low-competition. Between-participants competition decoding accuracy was not significant at any time window (Figure 18) for the *Episodic Animal Family Classifier*. However, a marginally significant between-participants classification accuracy (p = 0.097) of 61.32% was achieved in the 100 to 150 ms post-stimulus onset time window. This time window was used as the main window for

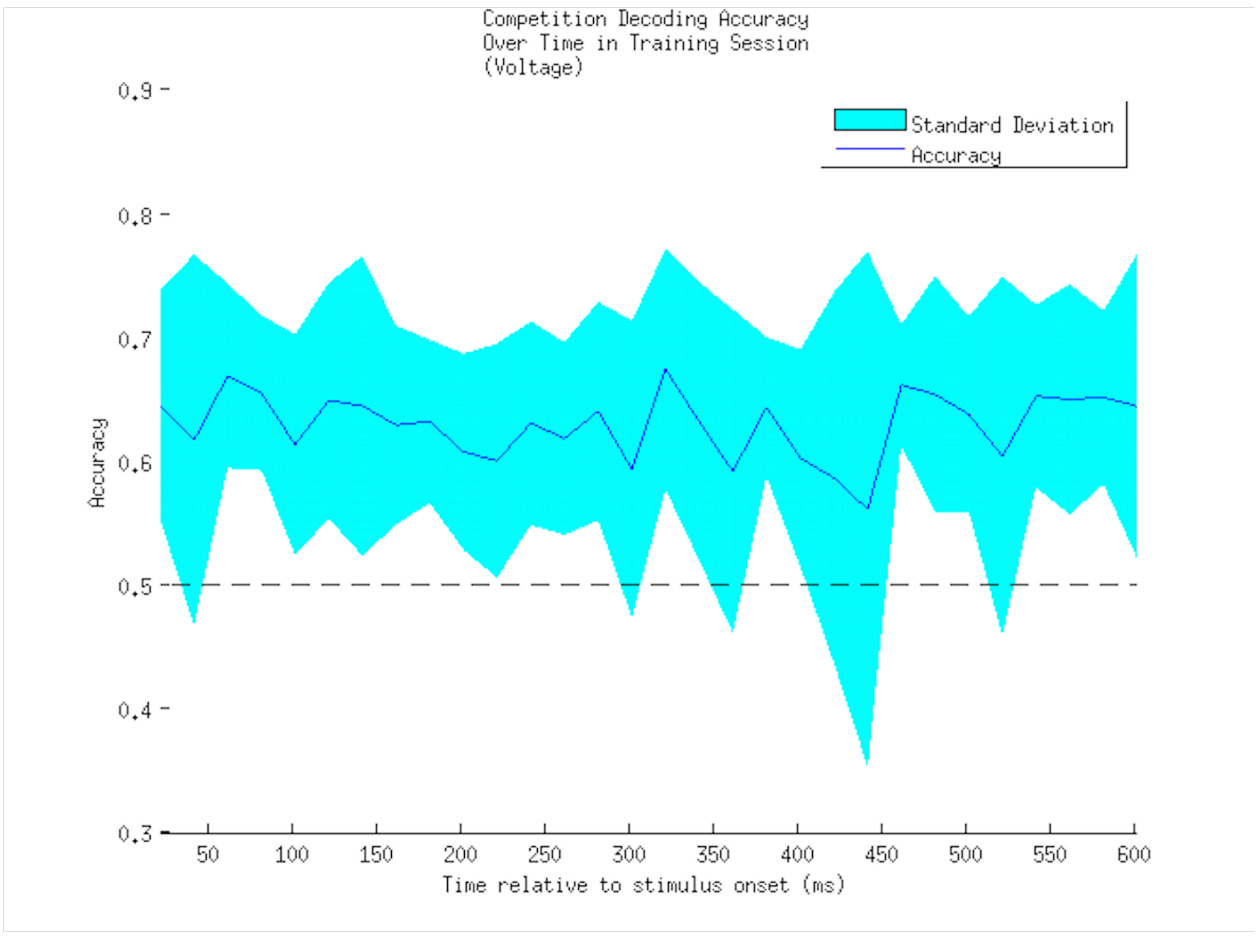


Figure 18. Competition decoding accuracy across pooled participants for the *Episodic Animal Family Classifier.* Significance determined by a within-participants permutation test with 100 permutations, as outlined in Rafidi et al. (2018). Marginally significant (p = 0.097) between-participants classification accuracy of 61.32% achieved when averaging voltages in the 100 to 150 ms post-stimulus window.

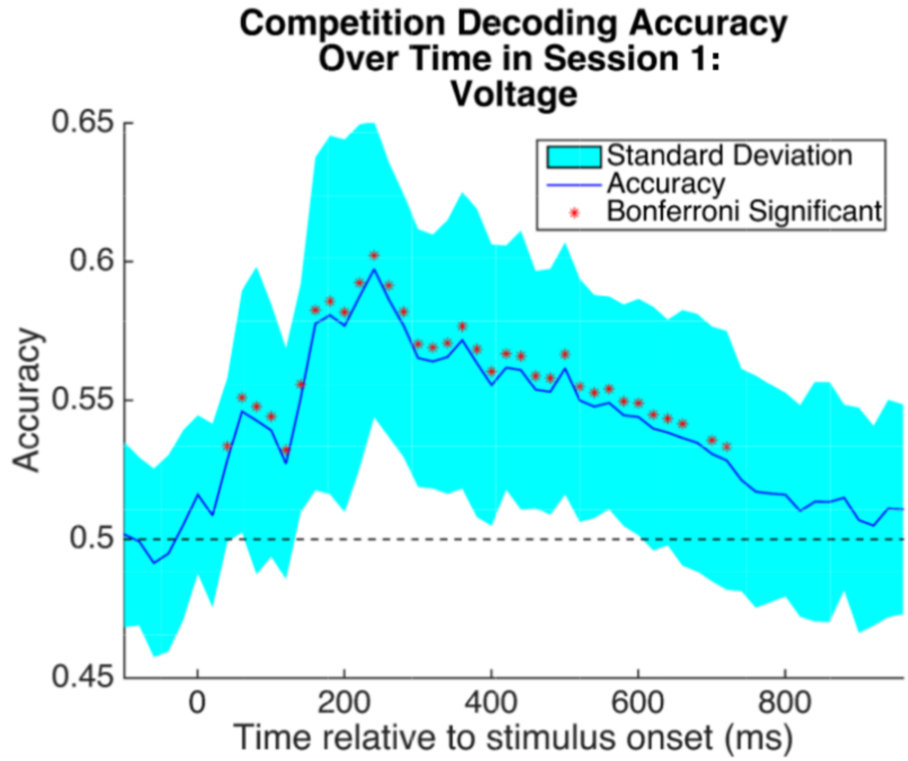


Figure 19. Results from Rafidi et al. (2018). Competition decoding accuracy aggregating all participants across each time window for *Semantic Category-Word Classifier*. Figure adapted from Rafidi et al. (2018).

subsequent analyses, as it provided the greatest competition decoding accuracy. Competition decoding accuracy from Rafidi et al. (2018) is displayed in Figure 19 for comparison.

Individual cross-validation classification decoding accuracies at the 100 ms post-stimulus time window are displayed in Figure 20. Classification decoding accuracy for all participants across five cross-validation folds at the 100 ms post-stimulus time window is shown in Figure 21. Aggregating classifier accuracy across participants smooths out the predictions. Average classifier accuracy for each participant, obtained by collapsing across folds, is displayed in Figure 22.

Long-Term Memory Predictions. Feeding EEG trials from the *Show and Tell Task* into *Episodic Animal Family Classifiers* developed specifically for each participant generated

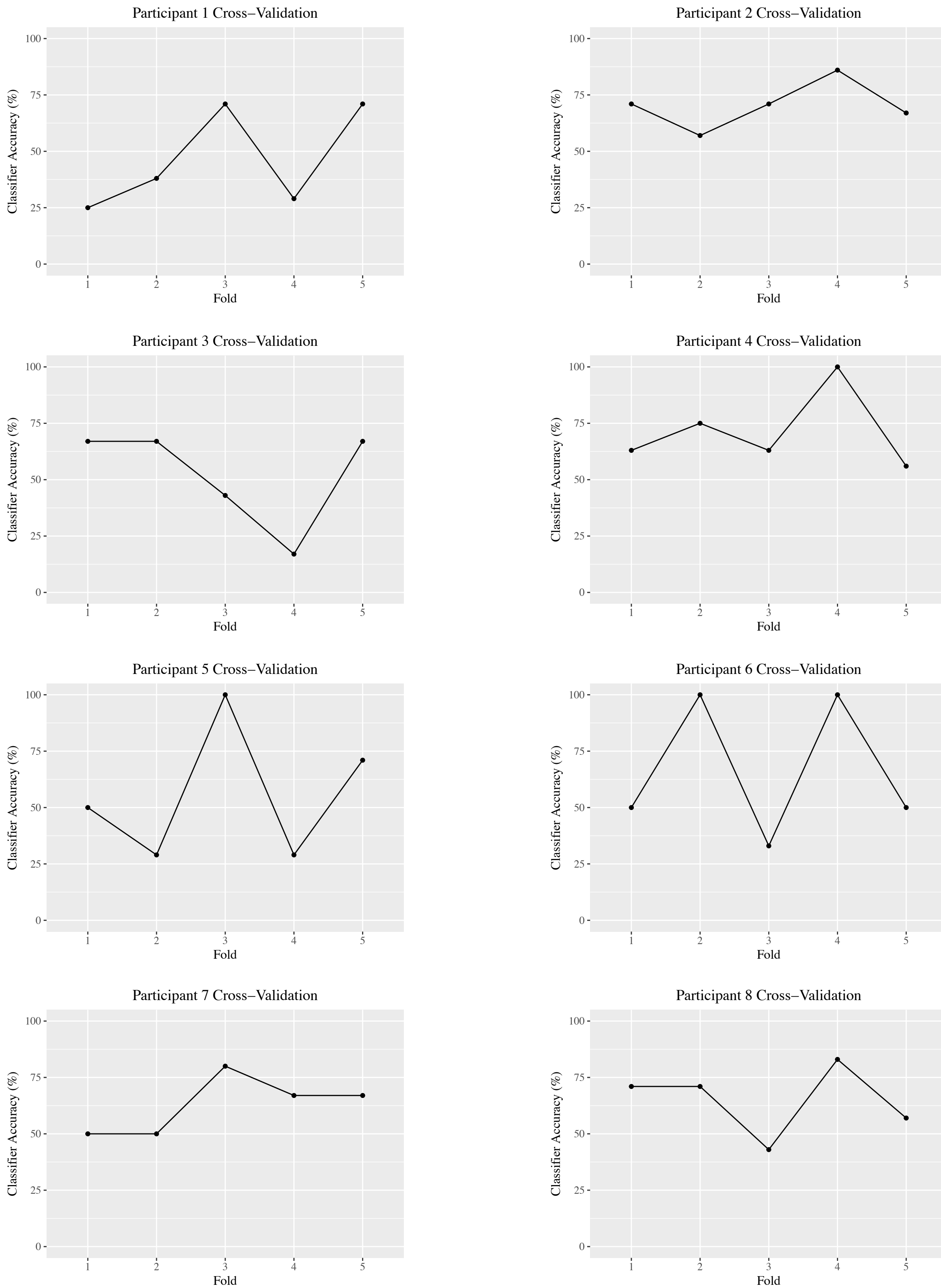


Figure 20. Within-participant classifier accuracy over five cross-validation folds for the *Episodic Animal Family Classifier*.

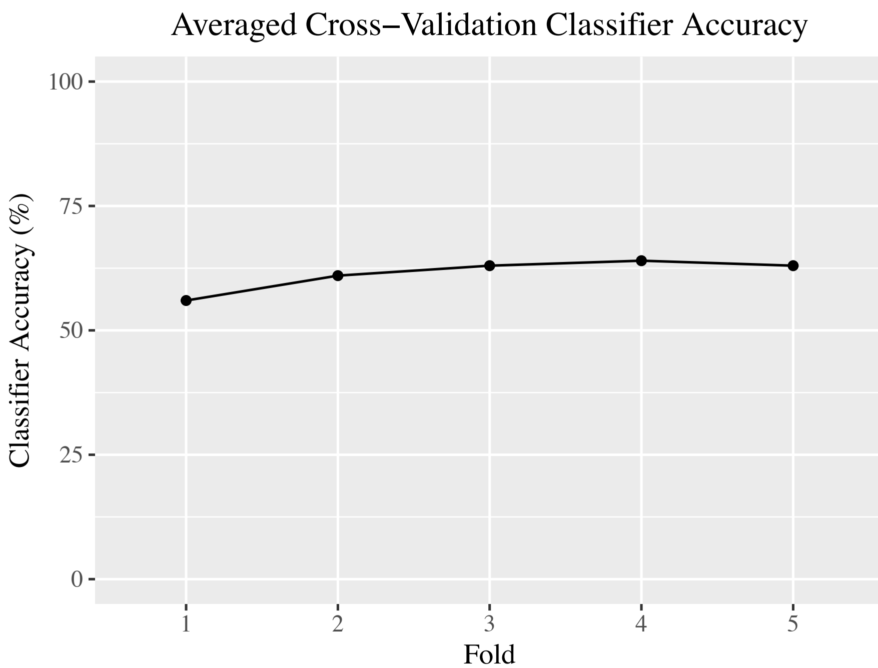


Figure 21. Average classifier accuracy over five cross-validation folds for the *Episodic Animal Family Classifier*. Obtained by collapsing across participants.

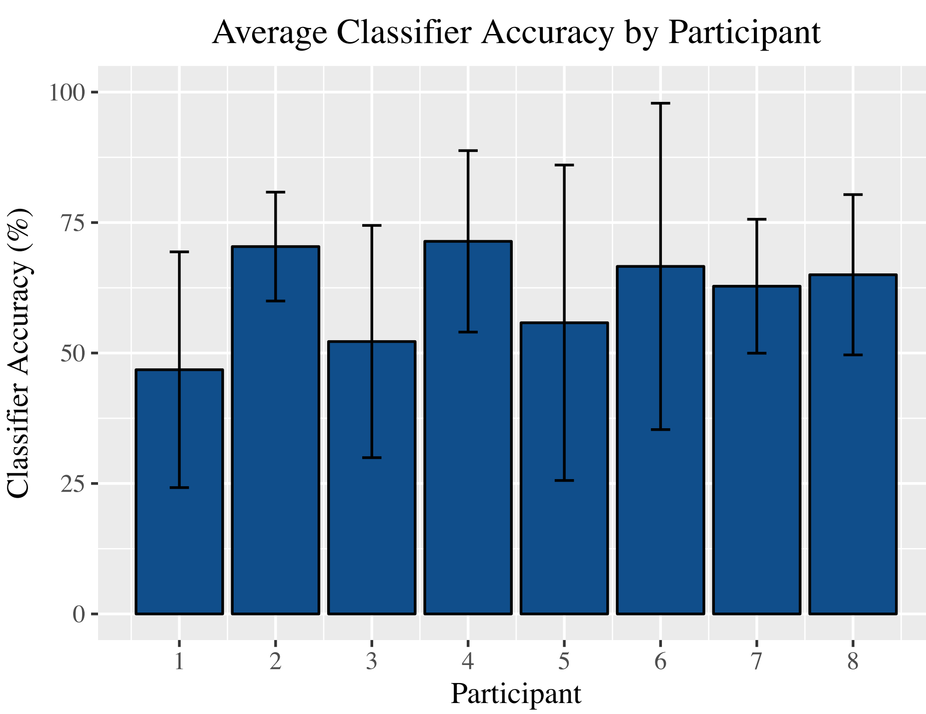


Figure 22. Average classifier accuracy by participant for the *Episodic Animal Family Classifier*. Obtained by collapsing across five cross-validation folds for each participant. Error bars = SD.

predictions about how many items would be remembered one week after acquisition. The classifiers produced values between 0 and 100. A classifier output between 0 and 0.05 meant that the classifier predicted that the EEG voltages from the input trial were low-competition signals. A classifier output between 0.95 and 1 meant that the classifier predicted that the EEG voltages from the input trial were high-competition signals.

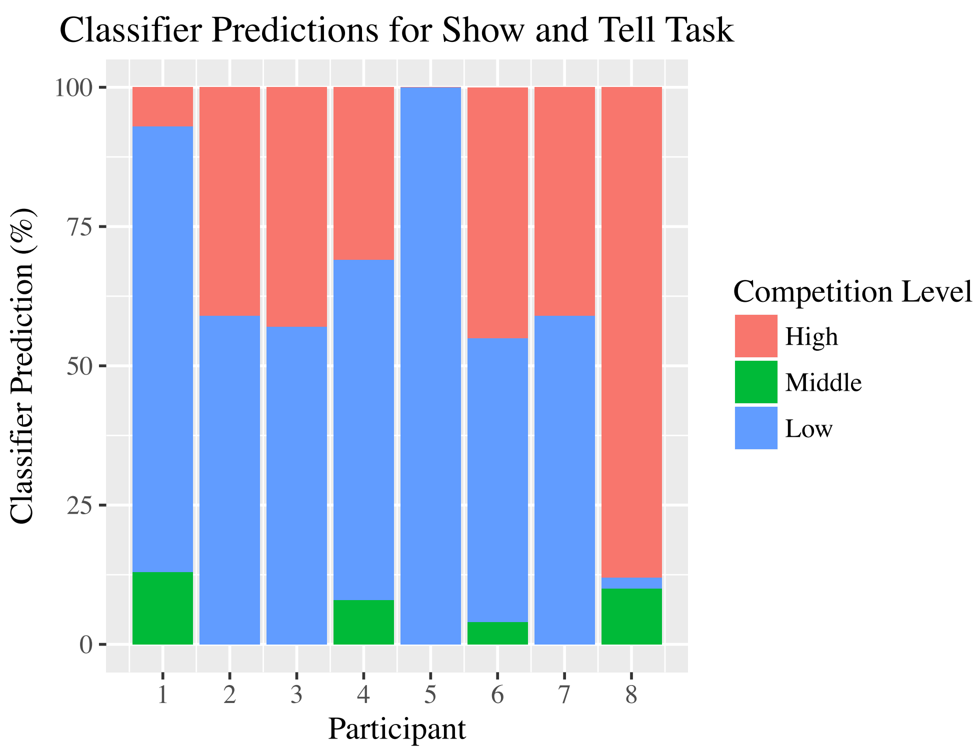


Figure 23. Episodic Animal Family Classifier predictions for *Show and Tell Task* final recall. Predictions, in the form of percentages, compared across inequivalent amounts of trials, for high- and low-competition, and for predictions that fell in-between high- and low-competition.

Because there was no retrieval-test phase in the *Show and Tell Task*, there was no single block during which every EEG trial could be classified once, as in the *Episodic Animal Family Classifier* training procedure. Additionally, no individual trial identifiers were retained, so individual items could not be tracked across retrieval-practice phases and final recall. Moreover, the amount of completed trials was not equivalent across participants, as some participants required more retrieval-practice training than others. In light of these limitations, classification predictions were made by aggregating all retrieval-practice trials that participants completed during the *Show and Tell Task,* and feeding all retrieval-practice trials into the *Episodic Animal Family Classifier* (Figure 23).

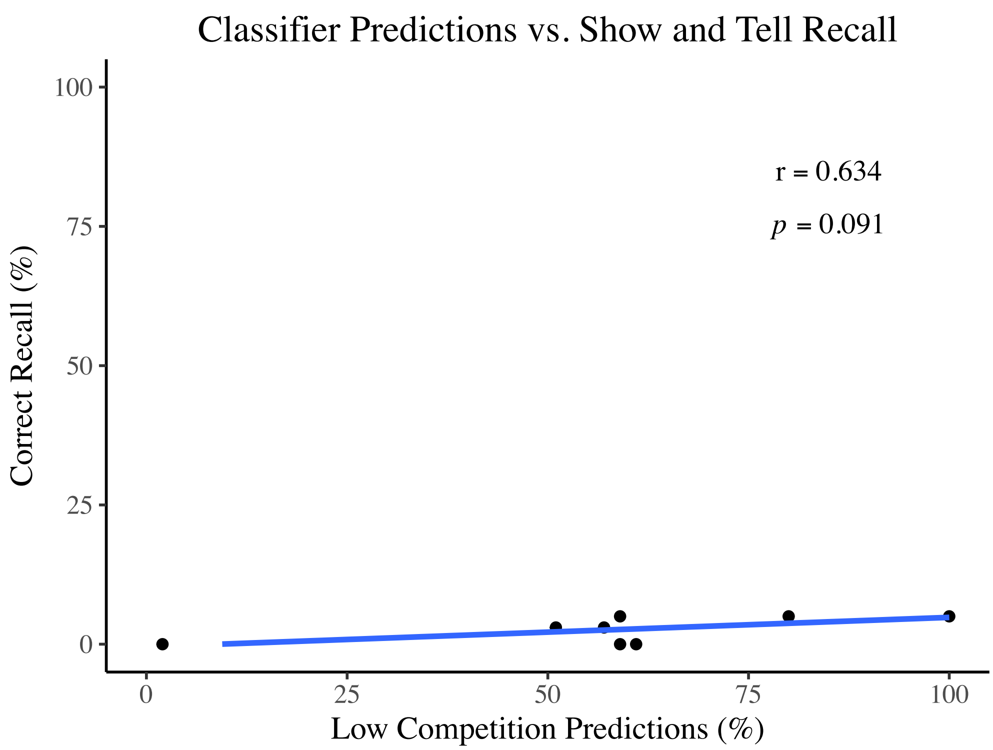


Figure 24. Episodic Animal Family Classifier predictions for percentage of low-competition trials compared to the actual percentage of trials that participants answered correctly at recall after a one-week delay. A marginally significant ($p = 0.091$) correlation is displayed, r = 0.634.

Pearson's product-moment correlation test revealed a marginally significant medium-strength correlation ($p = 0.091$) between *Episodic Animal Family Classifier* predictions of low-competition in the *Show and Tell Trials* and the actual percentage of trials that participants answered correctly at the final *Show and Tell* cued-recall test, r = 0.634 (Figure 24). This correlation co-efficient resembles the significance and classification accuracy (61.32%, *p* =

0.097) obtained by running cross-fold validation on the *Episodic Animal Family Classifier* training procedure trials, suggesting that the classifier is well adept to predict performance on the *Show and Tell Task*, as predicted. To see whether JOLs were a better predictor of final *Show and Tell* recall than the *Episodic Animal Family Classifier*, the relationship between JOLs and final *Show and Tell* recall was considered in comparison to the results of the correlation between *Episodic Animal Family Classifier* predictions and percent correct *Show and Tell* recall. Pearson's product-moment correlation test showed an insignificant negative correlation (p = 0.406) between between one-week delay JOLs and percent correct *Show and Tell* final recall, r = -0.343 (Figure 25).

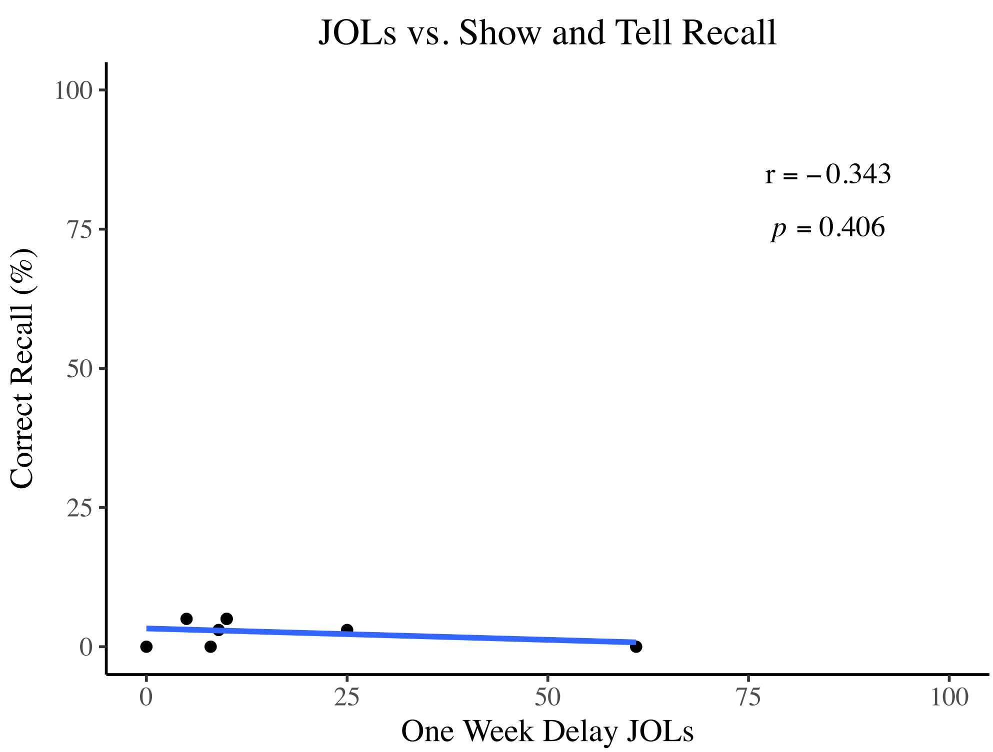


Figure 25. Judgments of learning (JOL) predictions for what percentage of *Show and Tell* items would be remembered after a one-week delay, compared to the percentage of correct recall on the final *Show and Tell* test. An insignificant (p = 0.406) is displayed, r = -0.343.

Discussion

The results and methodology from this experiment highlight the potential for a refined classification of memory competition. Despite the floor performance on the *Show and Tell Task* after the one week-delay, in addition to the limitations of not being able to run all replication analyses and compare competition decoding accuracies across classifiers and classifier testing procedure tasks, the results from the *Episodic Animal Family Classifier* suggest that memory competition signals in the brain can be effectively classified with limited resources. Further limitations, refinements, future directions and applications for this work are henceforth considered.

Significance of Results. Rafidi et al. (2018) achieved 64% classification decoding accuracy at $p < 0.0001$, compared to the *Episodic Animal Family Classifier* competition decoding accuracy of 61% at $p = 0.097$. Though competition decoding accuracy for the *Episodic Animal Family Classifier* is only marginally significant, the trend towards significance is noteworthy considering the technology used and methodological constraints.

Rafidi et al. (2018) used a BioSemi ActiveTwo EEG data acquisition device with 64-channels and a sampling rate of 512 Hz. The BioSemi system uses active electrodes that reduce noise and allow for a cleaner signal acquisition. Their system was additionally protected with a RF/EMI Faraday cage that filters out electrical noise in the environment. Given their setup, these researchers were poised to find clean signals.

Though it has been verified as a research device in the literature and its use is imperative for the translational aspect of this research, the Emotiv EPOC+ EEG headset produces noisier data than the BioSemi system for several reasons. First, the Emotiv EPOC+ is intended as a consumer-grade portable EEG device, while the BioSemi system is a research-grade device. Therefore, the Emotiv EPOC+ tends to produce noisier data and does not have any active noise filtering features. Second, the Emotiv EPOC+ is a one-size-fits-all headset, meaning that the electrodes are not placed in the exact same location on each participant's scalp. While efforts are made to be consistent in the placement of the headset across participants, localization of activity is more difficult to determine with such a device. Finally, the 64 electrode channels and the 512 Hz sampling rate of the BioSemi system loom in comparison to the 16 electrode channels and 128 Hz sampling rate of the Emotiv EPOC+. A lower sampling rate results in data with less acuity, making it potentially more difficult to pick up on relevant signals.

Considering the stark disadvantage that the Emotiv EPOC+ has in comparison to the BioSemi system, the marginally significant competition decoding accuracy of 61% suggests that the signals being observed are robust. Therefore, the classification accuracy was likely a result of the experimental design truly manipulating competition. Refining the procedure in order to make the classification accuracy significantly greater than chance performance is a logical direction for this work.

Refining Classifier Training and Testing Procedures. Redesigning several areas in the experiment could increase classifier accuracy, feasibility for brain-computer interface (BCI) applications, and participant comfort in tandem. A concern shared by all participants was the discomfort of the Emotiv EPOC+ headset; no objective measure was recorded about the comfortability of the headset, however all participants voiced their concerns during experiment sessions. These concerns were typically voiced after the participant had worn the headset for an extended period of time, usually more than an hour. Decreasing amount of time spent in the headset should be a priority for future iterations of this experiment, especially in the context of a BCI, as the discomfort that participants experienced could affect attention to tasks, thereby altering encoding of material. When there are not several training paradigms being compared, as there are in the current experiment, the time spent in the headset will naturally be shorter.

Participants also noted that it was easier to learn words in the *Swahili/English Task* and in the *Semantic Category-Word Classifier* training procedure than it was to learn the names and images for the *Episodic Animal Family Classifier* training procedure and the *Show and Tell Task.* Independently, they voiced that it was due to their relative unfamiliarity with the names used in the tasks. A subset of four participants were polled with supplementary questions in the Post-Experiment Questionnaire (Appendix 5) about their subjective experiences of task difficulty. Participants rated individual task difficulty on a five-point Likert scale from Very Easy to Very Difficult for the *Episodic Animal Family Classifier* training procedure (M = 3), the *Semantic Category-Word Classifier* training procedure (M = 1.75), the *Swahili/English Task* (M = 4), and the *Show and Tell Task* (M = 4). All four participants rated the *Episodic Animal Family Classifier* training procedure as being more difficult to learn than the *Semantic Category-Word Classifier.* They cited reasons such as, "I was often already familiar with most of the cat/item words, and there was actual semantic reason connecting them" and "All the animal names sound very similar, whereas the category/item pairs do not require as much memorization." Further, the subset of participants was divided on which classifier testing procedure was more difficult. While it may be a valid concern that the names are too unfamiliar, to the point where participants get caught up and distracted from the task by trying to learn them, making changes to better the learning procedures should take precedence over making the learning material easier if the goal of the research is to improve learning practices.

Since the names used in the *Episodic Animal Family Classifier* and the *Show and Tell Task* were so unfamiliar, it put them all on the same playing field in terms of difficulty to learn. This reflects how difficult it is to learn something completely new when there is nothing semantically familiar upon which to grasp. This re-emphasizes the intent of the experimental design and method for classifying competition. Since the *Semantic Category-Word Classifier* is trained using already established associations, as noted by the participants in the Post-Experiment Questionnaire, any competition detected during training phases does not reflect new learning. In contrast, the *Episodic Animal Family Classifier* introduced completely novel associations and tracked competition as participants learned these associations, which was a difficult task that introduced a lot of competition. When learning proves to be difficult, more time with the material is required to help resolve any residual competition.

Therefore, instead of making the *Episodic Animal Family Classifier* training procedure easier, participants should complete more rounds of retrieval-practice in future iterations of this experiment. In the *Show and Tell Task*, all participants completed at least one round of retrieval-practice; additional rounds only occurred if participants guessed an item incorrectly or did not respond in time during the retrieval-practice phase of the task. When completing the *Swahili/English Task*, all participants went through four rounds of retrieval-practice, which falls in line with the typical retrieval-practice procedures outlined in the Introduction.

Having a single retrieval-practice round in the *Episodic Animal Family Classifier* and the *Show and Tell Task* does not reflect an effective learning schedule, which would typically require multiple rounds of studying and testing, as was done in the *Swahili/English Task* (Karpicke & Roediger, 2008). As such, it seems reasonable that participants typically had greater recall for the *Swahili/English Task* one week post-learning. Increasing the amount of retrieval-practice trials would be a feasible change to the procedure if the Rafidi et al. (2018) replication was not run during the same sessions, making the time in the Emotiv headset approximately two hours per session. Ultimately, adding more rounds of retrieval-practice to the *Episodic Animal Family Classifier* training procedure and the *Show and Tell Task* would likely increase recall after a one-week delay, raise JOL predictions, expand the amount of trials that could be used for classification, and make neural signals of competition more distinguished, thereby improving classification accuracy.

Response Thresholds. People have different response thresholds that guide their probability of providing an answer when prompted with a question. While some participants in this experiment provided answers for all questions in the *Show and Tell Task* final test, some provided answers for only the ones with which they felt confident. Performance was approximately the same across participants, despite differences in the overall magnitude of responses for a given participant.

Response thresholds may have guided participants' strategies as they performed the tasks, steering some participants to respond only when they had enough information to confidently provide an answer. Several participants verbally noted that they would wait for the second letter of the name for an animal or preschooler, which was unique to that item, to appear before they guessed the full name. Participants noted that they felt that the images were too difficult to remember. Future iterations of this methodology should lower the threshold for a failed trial during the slow-reveal of letters in the retrieval-practice phase of the *Episodic Animal Family Classifier* training procedure and the *Show and Tell Task*. High-competition trials should be marked as unlearned if participants do not respond after two letters are revealed. Low-competition trials should be marked as unlearned if they do not respond when given only the first letter. This will increase the training difference between types of trials, ideally drawing a more defined line between high- and low-competition.

Controlling for Difficulty. An initial concern when developing the *Episodic Animal Family Classifier* training procedure was how the results would differentiate between levels of competition and task difficulty. To a large extent, this experiment confounds the two. Conflating these factors is not necessarily harming the attempts to focus on classifying competition, as much of what the difficulty of a task measures is in fact competition. When something is more difficult, an imbalance of competition is always created. The *Episodic Animal Family Classifier* was trained using material that is endowed with many different facets of competition (Figure 9). Having so much competitive material increases chances of accurately classifying competition, but it admittedly makes the tasks more difficult.

In the *Episodic Animal Family Classifier* training procedure in its current form, there are more high-competition trials than there are low-competition trials. Attempting to equalize the number of high- and low-competition trials should be a consideration for future experiments. This could be achieved by introducing a greater number of low-competition animal families (though, this would create a difference in total number of animal families to be learned) or by increasing the amount of retrieval-practice trials for low-competition animals, relative to high-competition animals, equalizing the total number of trials across competition levels despite there being fewer low-competition animals overall.

In future experiments, a secondary task could be added in attempt to directly isolate difficulty from competition. For example, every animal group in the *Episodic Animal Family Classifier* training procedure could have six members, making all of the animals high-competition animals. In half of the retrieval-practice trials, participants would be tasked with remembering a six-digit number. A secondary task such as this would make retrieving the correct animal name more difficult, while establishing a consistent level of high-competition across all animals. In this example, however, there is still competition for memory resources, albeit different types of memory (working memory vs. long-term memory). It would be interesting to see if competition in different types of memory could exist independent of one another, or if overloading working memory competition would affect long-term memory and retrieval competition.

Previous work has shown that blink rate increases as a function of task difficulty (Tanaka & Yamaoka, 2011). As an added manipulation check for difficulty, electrode sensors could be placed on participants' faces to gauge blink rate during difficult trials or high-competition trials compared to not-difficult trials or low-competition trials.

Future Directions and Applications. The continued development of reliable EEG, making it both more portable and affordable, puts powerful technology in the hands of more users, opening the door for a new era of education technology. Specifically, locating and harnessing robust neural signals that predict long-term retention provides a unique opportunity to create innovative and adaptive study strategies for the modern learner. Future work will assess the efficacy of an EEG-based neurofeedback BCI that is designed to guide students and everyday learners to adapt their study habits to learn more efficiently and with greater long-term retention.

Using neurofeedback technology, learners can self-regulate their cognitive performance by gaining an accurate understanding of their own brain states (Sitaram et al., 2016). Figure 26 displays a basic outline of a closed-loop, neurofeedback BCI. As learners study new information, classifiers analyze EEG data in real-time in order to get a live readout of competitive brain states. The BCI is then able to adjust the future study schedule, allocating additional study time for items that are still in a labile, high-competition state.

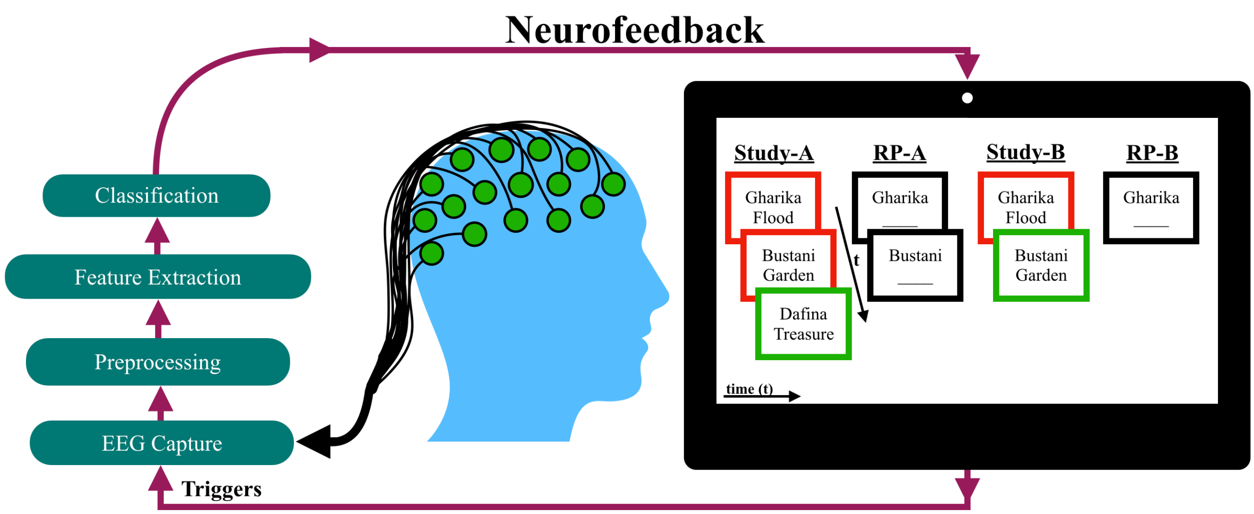


Figure 26. Neurofeedback Brain-Computer Interface (BCI). Green and red boxes indicate word pairs classified as displaying low- and high-levels of competition, respectively.

An additional advantage that the *Episodic Animal Family Classifier* holds over the classifier used in Rafidi et al. (2018) is that its long-term memory predictions are based on instantaneous readings of competition. The classifier in Rafidi et al. (2018) calculated predictions by taking the competition reading for a given trial at several time points, and computing the difference in competition between the beginning and end of the training session. While this yielded significant results and suggested that competition could be accurately tracked in the brain, this "competition drop score" is not conducive to BCI neurofeedback applications that would ideally take trial-by-trial readings of competition and adjust subsequent study presentation orders. The ability of the *Episodic Animal Family Classifier* to generate these predictions, given a single trial, is ideal for BCI. Future work should explore the competition decoding accuracy of the *Episodic Animal Family Classifier* in feature spaces other than voltages (e.g. frequency bands and specific spatial locations), and consider how a refined classifier can best be employed in a neurofeedback BCI paradigm.

References

Anderson, J. R., & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General, 128*(2), 186-197. doi:10.1037/0096-3445.128.2.186

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(5), 1063-1087. doi:10.1037//0278-7393.20.5.1063

Badcock, N. A., Mousikou, P., Mahajan, Y., de Lissa, P., Thie, J., & McArthur, G. (2013). Validation of the Emotiv EPOC ® EEG gaming system for measuring research quality auditory ERPs. PeerJ, 1, e38. https://doi.org/10.7717/peerj.38

Badcock, N. A., Preece, K. A., de Wit, B., Glenn, K., Fieder, N., Thie, J., & McArthur, G. (2015). Validation of the Emotiv EPOC EEG system for research quality auditory event-related potentials in children. PeerJ, 3, e907. https://doi.org/10.7717/peerj.907

Broek, G. V., Takashima, A., Wiklund-Hörnqvist, C., Wirebring, L. K., Segers, E., Verhoeven, L., & Nyberg, L. (2016). Neurocognitive mechanisms of the "testing effect": A review. *Trends in Neuroscience and Education, 5*(2), 52-66. doi:10.1016/j.tine.2016.05.001

Carpenter, S. K., & Delosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition, 34*(2), 268-276. doi:10.3758/bf03193405

Collins, B.(1999). Questions About Angels. Pittsburgh: University of Pittsburgh Press. Retrieved from Project MUSE database.

DeSouza, J. F. X. (2014). EEG Tutorial. Retrieved from https://bitbucket.org/joelab/eeg-tutorial/wiki/Home

Ebbinghaus, H. (1895). Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie.

Ebbinghaus, H. (1913). Memory: A contribution to experimental psychology (H. A. Ruger & C. E. Bussenius, Trans.).

Ekandem, J. I., Davis, T. A., Alvarez, I., James, M. T., & Gilbert, J. E. (2012). Evaluating the ergonomics of BCI devices for research and experimentation. Ergonomics, 55(5), 592–598. https://doi.org/10.1080/00140139.2012.662527

Hirschstein, Z. (2018). Towards Improving Learning with Consumer-Grade, Closed-Loop, Electroencephalographic Neurofeedback. *Bard Digital Commons*.

Hulbert, J. C., & Norman, K. A. (2014). Neural Differentiation Tracks Improved Recall of Competing Memories Following Interleaved Study and Retrieval-practice. *Cerebral Cortex, 25*(10), 3994-4008. doi:10.1093/cercor/bhu284

Karpicke, J. D., Butler, A. C., & Iii, H. L. (2009). Metacognitive strategies in student learning: Do students practice retrieval when they study on their own? *Memory, 17*(4), 471-479. doi:10.1080/09658210802647009

Karpicke, J. D., & Roediger, H. L. (2008). The Critical Importance of Retrieval for Learning. *Science, 319*(5865), 966-968. doi:10.1126/science.1152408

Luck, S. J. (2014) An Introduction to Event-Related Potential Technique, Second Edition. Cambridge, MA: MIT Press.

Manning, J. (2019). Episodic memory: mental time travel or a quantum "memory wave" function? PsyArXiv, 1–15. Retrieved from https://psyarxiv.com/6zjwb/download

Maskeliunas, R., Damasevicius, R., Martisius, I., & Vasiljevas, M. (2016). Consumer grade EEG devices: are they usable for control tasks? PeerJ, 4, e1746. https://doi.org/10.7717/peerj.

1746

Murayama, K., Miyatsu, T., Buchli, D., & Storm, B. C. (2014). Forgetting as a consequence of retrieval: A meta-analytic review of retrieval-induced forgetting. *Psychological Bulletin, 140*(5), 1383-1409. doi:10.1037/a0037505

Noh, E., Herzmann, G., Curran, T., & Sa, V. R. (2013). Using single-trial EEG to predict and analyze subsequent memory. *NeuroImage, 84*, 712-723. doi:10.1016/j.neuroimage. 2013.09.028

Norman, K. A., Newman, E., Detre, G., & Polyn, S. (2006). How Inhibitory Oscillations Can Train Neural Networks and Punish Competitors. *Neural Computation, 18*(7), 1577-1610. doi:10.1162/neco.2006.18.7.1577

Oostenveld, R., Fries, P., Maris, E., Schoffelen, JM (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, Article ID 156869, doi:10.1155/2011/156869

Paller, K. A., Kutas, M., & Mayes, A. R. (1987). Neural correlates of encoding in an incidental learning paradigm. *Electroencephalography and Clinical Neurophysiology, 67*(4), 360-371. doi:10.1016/0013-4694(87)90124-6

Putnam, A. L., Sungkhasettee, V. W., & Roediger, H. L. (2016). Optimizing Learning in College. *Perspectives on Psychological Science, 11*(5), 652-660. doi:10.1177/1745691616645770

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rafidi, N. S. (2012). *A Brain-Computer Interface for Enhanced Learning Using Classification of*

EEG Data. Retrieved from Princeton University Undergraduate Senior Theses. (http://*arks.princeton.edu/ark:/88435/dsp01rf55z972n*)

Rafidi, N. S., Hulbert, J. C., Pacheco, P., & Norman, K. A. (2018). Reductions in Retrieval Competition Predict the Benefit of Repeated Testing. *Scientific Reports, 8*(1), doi: 10.1101/292557

Rodriguez, F., Kataoka, S., Rivas, M. J., Kadandale, P., Nili, A., & Warschauer, M. (2018). Do spacing and self-testing predict learning outcomes? *Active Learning in Higher Education,* 146978741877418. doi:10.1177/1469787418774185

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, 140(6), 1432-1463. doi:10.1037/a0037559

Sanquist, T. F., Rohrbaugh, J. W., Syndulko, K., & Lindsley, D. B. (1980). Electrocortical Signs of Levels of Processing: Perceptual Analysis and Recognition Memory. *Psychophysiology*, 17(6), 568-576. doi:10.1111/j.1469-8986.1980.tb02299.x

Schacter, D. (2002). The Seven Sins of Memory: Insights from Psychology and Cognitive Neuroscience. Foundations in Social Neuroscience, 54(3), 113–147.

Sitaram, R., Ros, T., Stoeckel, L., Haller, S., Scharnowski, F., Lewis-Peacock, J., … Sulzer, J. (2017). Closed-loop brain training: The science of neurofeedback. Nature Reviews Neuroscience, 18(2), 86–100. https://doi.org/10.1038/nrn.2016.164

Tanaka, Y., & Yamaoka, K. (2011). Blink Activity and Task Difficulty. Perceptual and Motor Skills, 77(1), 55–66. https://doi.org/10.2466/pms.1993.77.1.55

Tank, S. (2013). A brain-computer interface for enhanced learning using classification of EEG data. Princeton University.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods*, 42(3), 671-684.

Xian, Y., Schiele, B., & Akata, Z. (2017). Zero-shot learning - The good, the bad and the ugly. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition*, CVPR 2017, 2017–January, 3077–3086. https://doi.org/10.1109/CVPR.2017.328

Appendices

1. Swahili/English task stimuli
2. Category-Word Classifier stimuli
3. Distractor Task: Sudoku
4. Demographics Questionnaire
5. Post-Experiment Questionnaire
6. IRB Confirmation
7. Accepted IRB Protocol

A. Semantic Category-Word Classifier Stimuli

ALCOHOL - BRANDY
ALCOHOL - CHAMPAGNE
ALCOHOL - GIN
ALCOHOL - SCOTCH
ALCOHOL - TEQUILA
ALCOHOL - VODKA
ALCOHOL - WINE
ALCOHOL - WHISKEY
APPLIANCE - BLENDER
APPLIANCE - DRYER
APPLIANCE - JUICER
APPLIANCE - MICROWAVE
APPLIANCE - REFRIGERATOR
APPLIANCE - STOVE
APPLIANCE - TOASTER
APPLIANCE - WASHER
BUILDING - CHURCH
BUILDING - MONUMENT
BUILDING - MUSEUM
BUILDING - RESTAURANT
BUILDING - SCHOOL
BUILDING - SKYSCRAPER
BUILDING - STADIUM
BUILDING - TOWER
CANDY - BUTTERSCOTCH
CANDY - CARAMEL
CANDY - GUM
CANDY - LICORICE
CANDY - LOLLIPOP
CANDY - MINT
CANDY - SUCKER
CANDY - TAFFY
DANCE - DISCO
DANCE - FOLK
DANCE - JIG
DANCE - POLKA
DANCE - SQUARE
DANCE - TAP
DANCE - TWIST
DANCE - WALTZ
DISEASE - DIABETES
DISEASE - HEPATITIS
DISEASE - MEASLES
DISEASE - MUMPS
DISEASE - PNEUMONIA
DISEASE - POLIO
DISEASE - SMALLPOX
DISEASE - TYPHOID
ELEMENT - ARGON
ELEMENT - BARIUM
ELEMENT - IODINE
ELEMENT - MERCURY
ELEMENT - RADIUM
ELEMENT - SULFUR
ELEMENT - URANIUM
ELEMENT - ZINC
EMOTION - ENVY
EMOTION - EXCITEMENT
EMOTION - FEAR
EMOTION - LOVE
EMOTION - PITY
EMOTION - SHAME
EMOTION - SORROW
EMOTION - TENSION
FABRIC - BURLAP
FABRIC - DENIM
FABRIC - FLANNEL
FABRIC - HEMP
FABRIC - LACE
FABRIC - NYLON
FABRIC - SATIN
FABRIC - VELVET
FLOWER - DAISY
FLOWER - IRIS
FLOWER - LILY
FLOWER - ORCHID
FLOWER - PANSY
FLOWER - POPPY
FLOWER - TULIP
FLOWER - VIOLET
GEM - AMETHYST
GEM - JADE
GEM - ONYX
GEM - OPAL
GEM - QUARTZ
GEM - SAPPHIRE
GEM - TOPAZ
GEM - TURQUOISE
INSECT - BEETLE
INSECT - ANT
INSECT - CATERPILLAR
INSECT - GRASSHOPPER
INSECT - LADYBUG
INSECT - MOSQUITO
INSECT - TICK
INSECT - WASP
LANDFORM - CRATER
LANDFORM - DUNE
LANDFORM - GORGE
LANDFORM - GULLY
LANDFORM - ISLAND
LANDFORM - PLATEAU
LANDFORM - RAVINE
LANDFORM - RIDGE
PROFESSION - ARTIST
PROFESSION - DENTIST
PROFESSION - ELECTRICIAN
PROFESSION - JOURNALIST
PROFESSION - PILOT
PROFESSION - SECRETARY
PROFESSION - TEACHER
PROFESSION - WRITER
SEASONING - BASIL
SEASONING - CLOVE
SEASONING - CUMIN
SEASONING - DILL
SEASONING - GINGER
SEASONING - NUTMEG
SEASONING - OREGANO
SEASONING - PAPRIKA
TOOL - DRILL
TOOL - LADDER
TOOL - LEVEL
TOOL - PLIERS
TOOL - SCREWDRIVER
TOOL - SHOVEL
TOOL - VISE
TOOL - WRENCH
TREE - ASPEN
TREE - BIRCH
TREE - CEDAR
TREE - EVERGREEN
TREE - HOLLY
TREE - PINE
TREE - SPRUCE
TREE - WILLOW
VEGETABLE - ASPARAGUS
VEGETABLE - BROCCOLI
VEGETABLE - EGGPLANT
VEGETABLE - ONION
VEGETABLE - POTATO
VEGETABLE - RUTABAGA
VEGETABLE - SPINACH
VEGETABLE - ZUCCHINI
WEAPON - ARROW
WEAPON - BOMB
WEAPON - CLUB
WEAPON - DAGGER
WEAPON - POISON
WEAPON - ROCKET
WEAPON - SWORD
WEAPON - WHIP

B. Swahili/English Task Stimuli

BUU - MAGGOT
THELUJI - SNOW
FARASI - HORSE
MAITI - CORPSE
PUNDA - DONKEY
NYANYA - TOMATO
NDOO - BUCKET
MASHUA - BOAT
GOTI - KNEE
KAPUTULA - SHORTS
ZULIA - CARPET
FUNUNU - RUMOR
ELIMU - SCIENCE
BUSTANI - GARDEN
LESO - SCARF
DAFINA - TREASURE
ADHAMA - HONOR
TUMBILI - MONKEY
GODORO - MATTRESS
VUKE - STEAM
FAGIO - BROOM
SUMU - POISON
CHAKULA - FOOD
KABURI - GRAVE
PIPA - BARREL
TABIBU - DOCTOR
YAI - EGG
HARIRI - SILK
KAA - CRAB
MALKIA - QUEEN

PAZIA - CURTAIN
REMBO - ORNAMENT
POMBE - BEER
ADUI - ENEMY
ZIWA - LAKE
SALA - PRAYER
ROHO - SOUL
USINGIZI - SLEEP
EMBE - MANGO
ZEITUNI - OLIVES
WALI - RICE
MBWA - DOG
GARI - CAR
MLIMA - MOUNTAIN
SURA - FACE
SAMAKI - FISH
MLANGO - DOOR
KOFIA - HAT
MENDE - BEETLE
UPEPO - WIND
KOLEO - SHOVEL
SHAKA - PROBLEM
MTO - PILLOW
JESHI - ARMY
NJIA - ROAD
MAJI - WATER
NYOKA - SNAKE
RINDA - DRESS
UMBU - SIBLING
MAZIWA - MILK

C. Sudoku Puzzle Distractor Task

Puzzle #1 - E

7	3		2		6	4		
		2	9			5		6
9	4		1		7			
8	9	1		4			6	
		4		6			3	8
		7		9		1	4	
5		8		1		6		9
4					8	7		
	2		5				1	4

D. Participant Demographics Form

PARTICIPANT DEMOGRAPHICS FORM

Study Name:
Room Number:
Experimenter:
Experiment Date: __ __ / __ __ / __ __ __ __

1. Gender: ☐ Female ☐ Male ☐ Other: ______________
2. Handedness: ☐ Left ☐ Right ☐ Ambidextrous
3. Are you a native speaker of English?: ☐ Yes ☐ No
4. Date of Birth: __ __ / __ __ / __ __ __ __
5. Are you 18 years or older with normal/corrected-to-normal color vision and no history of a learning disability or attentional disorder? ☐ Yes ☐ No
6. Race ("X" ONLY one with which you MOST CLOSELY identify):

   ☐ American Indian or Alaska Native

   ☐ Asian

   ☐ Black or African American

   ☐ Native Hawaiian or Other Pacific Islander

   ☐ White

   ☐ More than one race

   ☐ Unknown or not reported

7. Ethnicity ("X" ONLY one with which you MOST CLOSELY identify):

   ☐ Hispanic or Latino

   ☐ Not Hispanic or Latino

   ☐ Unknown or not reported

This information is being collected in accordance with the National Institute of Health's policy of recording data about subject diversity and is not analyzed in this experiment. The above gender/race/ethnicity category labels were established by the NIH.

E. Post-Experiment Questionnaire

Post-Experiment Questionnaire

1. **Participant Number**

____

Post-Experiment Questionnaire

2. **Did you expect there to be a final test the next week for the Swahili words?**
Mark only one oval.

- ( ) Yes
- ( ) No
- ( ) Other: ______________________________

Post-Experiment Questionnaire

3. **Did you expect there to be a final test the next week for the preschooler items?**
Mark only one oval.

- ( ) Yes
- ( ) No
- ( ) Other: ______________________________

Post-Experiment Questionnaire

4. **What study strategies did you use to learn the Swahili words?**

____
____
____
____
____

Post-Experiment Questionnaire

5. **What study strategies did you use to learn the preschooler's items?**

Post-Experiment Questionnaire

6. **Did you remember as many items as you thought you would one week later?**
   Mark only one oval.

   - Yes
   - No
   - Other:

Post-Experiment Questionnaire

7. **Did you remember as many Swahili words as you thought you would one week later?**
   Mark only one oval.

   - Yes
   - No
   - Other:

Post-Experiment Questionnaire

8. **Were there any words, pictures, or names in the experiment that stood out as particularly meaningful to you?**

Post-Experiment Questionnaire

9. **How difficult was it to learn: Animal Names**

Mark only one oval.

	1	2	3	4	5	
Very easy	⬭	⬭	⬭	⬭	⬭	Very difficult

10. **How difficult was it to learn: Category/Item Pairs**

Mark only one oval.

	1	2	3	4	5	
Very easy	⬭	⬭	⬭	⬭	⬭	Very difficult

11. **How difficult was it to learn: Preschooler Items**

Mark only one oval.

	1	2	3	4	5	
Very easy	⬭	⬭	⬭	⬭	⬭	Very difficult

12. **How difficult was it to learn: Swahili Words**

Mark only one oval.

	1	2	3	4	5	
Very easy	⬭	⬭	⬭	⬭	⬭	Very difficult

Post-Experiment Questionnaire

13. **Which was more difficult to learn:**

Mark only one oval.

- ⬭ Animal Names
- ⬭ Category/Item Pairs

14. **Why was the option that you chose more difficult to learn? What made it difficult?**

____
____
____
____
____

Post-Experiment Questionnaire

15. **Which was more difficult to learn:**

Mark only one oval.

- ( ) Preschooler Items
- ( ) Swahili Words

16. **Why was the option that you chose more difficult to learn? What made it difficult?**

Post-Experiment Questionnaire

17. **What hypotheses do you think are being investigated in this experiment? In other words, what do you think this experiment is truly studying?**

Post-Experiment Questionnaire

Thank you for your participation. Please alert your experimenter that you have finished this questionnaire.

F. IRB Confirmation

Bard College   Institutional Review Board

Date: December 19, 2018

To: Noah Libby (nl8800@bard.edu)
Cc: Justin Hulbert (jhulbert@bard.edu)
From: Sanjay DeSilva, IRB Chair

Re: Paying Attention to Real-Time Neurofeedback

DECISION: APPROVED

Dear Noah,

The Bard Institutional Review Board reviewed your renewal request for the previously approved proposal 2018FEB07-HIR. Your proposal is approved through December 19, 2019.

Please notify the IRB if your methodology changes or unexpected events arise. We wish you the best of luck with your research.

We wish you the best of luck with your research.

Sanjay DeSilva
desilva@bard.edu
IRB Chair

PO Box 5000, Annandale-on-Hudson, New York 12504-5000 **Phone** 845-758-6822

SECTION 1

1. **Last name:** Libby
2. **First Name:** Noah
3. **E-mail:** nl8800@bard.edu
4. **Phone number:** 207-671-0971
5. **Academic program:** Psychology
6. **Status:** Student
7. **Name of faculty adviser/sponsor:** Justin Hulbert
8. **Adviser's/sponsor's e-mail:** jhulbert@bard.edu
9. **Today's date:** December 6, 2018

SECTION 2

1. **I have read the IRB's Categories of Review, and my proposal qualifies for a:** Renewal
2. **Do you have external funding for this research?** No
    a. **If so, state name of granting institution:** Not applicable
3. **Begin date:** Upon approval
4. **End date:** Ongoing, pending regular IRB reviews
5. **Title:** Paying Attention to Real-Time Neurofeedback

Research question: Can neurofeedback increase one's ability to learn? Our growing understanding of brain functioning, in combination with advanced computational techniques that can "read" the mind in near real time, promise sizeable advances in human potential. For example, the real-time information about one's brain state provided by neurofeedback has the potential to retrain brain dynamics disrupted due to stroke (Kober et al., 2015), as well as coax healthy individuals to adopt brain states associated with heightened attention (deBettencourt, Cohen, Lee, Norman, & Turk-Browne, 2015). In other words, feedback about the brain can help the brain train itself to adaptively adopt and adjust to the demands of an ever-changing and increasingly-digital world.

While a great deal of learning depends on focused attention, there are other neurocognitive factors that help determine what information is encoded and later made accessible for use. My Senior Project aims to develop a means of classifying the level of attentional focus, memory competition, and mnemonic engagement using brainwaves (electroencephalograms, or EEG) recorded non-invasively from participants' scalps. After validating the accuracy of these readings, I plan to test whether altering visual study materials based on classifier evidence (for the optimized attention/memory states) improves learning. This research intends to resolve questions about 1) how learning about one's brain states can affect behavior and 2) how viable real-time neurological feedback paradigms are at institutions such as Bard.

6. **Will your participants include individuals from specific populations (e.g., children, pregnant women, prisoners, or the cognitively impaired)?** No
7. **If your participants will include individuals from specific populations, please specify the population(s) and briefly describe any special precautions you will use.** Not applicable
8. **Briefly describe how you will recruit participants (e.g., Who will approach participants? What is the source of the participants?).** While future research in this area may focus on individuals with attentional/learning disabilities, participants recruited under this proposed protocol would be healthy adults who are free of diagnosed neurological/attentional/learning disabilities, between the ages of 18-35, and with normal/corrected-to-normal color vision. Participants additionally need to be willing/able to have felt electrode tips, electrode caps, Signa Gel and/or saline solution introduced to their hair/scalp and sit relatively still without excessive blinking throughout the experiment, in some cases while resting their chin on a comfortable platform to maintain a standard distance from the computer monitor; certain EEG and language-centric components will require participants to be right-handed and/or have been exposed to English regularly since early childhood/native English speakers. Participants will be drawn from Bard College and surrounding communities. Recruitment materials (posters, flyers, messages distributed via electronic bulletin boards/listservs/social media, and/or advertisements placed in local online/printed periodicals—see Appendix A) will direct interested parties to contact the researcher at nl8800@bard.edu or to potentially learn more/sign up for an appointment directly through the Psychology Program's online experiment booking system, https://bardresearch.sona-systems.com/. The booking site used would also host information about the study and allow interested members of the existing participant pool to sign up. On first contact, participants will be asked to confirm their eligibility for the particular study in question and their desire to participate. Following this, they would have the opportunity to schedule an appointment. Upon arrival at their scheduled appointment, participants will go through the informed consent process (see Appendix B for example language used in these materials). Depending on the length of their scheduled session, participants may be offered a token piece of candy and raffle entries for Amazon gift cards (ranging from $25-50), with winners selected at random by May 22, 2019, plus any bonuses introduced during the procedure. Certain phases may

offer participants bonus incentives, totaling up to $10 per participant per session, on the basis of their performance. Should a participant be invited back for additional experiment sessions, they will be compensated $5/hour for every hour in these additional sessions. At present, all mentions of "the research group" or "the researcher" simply refer to the investigator (Noah Libby), who will be leading the recruitment, testing, and analysis efforts, and his faculty supervisor, Justin Hulbert.

9. **Briefly describe the procedures you will be using to conduct your research. Include descriptions of what tasks your participants will be asked to do, and about how much time will be expected of each individual. NOTE: If you have supporting materials (recruitment posters, printed surveys, etc.) please email these documents separately as attachments to IRB@bard.edu. Name your attachments with your last name and a brief description (e.g., "WatsonConsentForm.doc").**
    a. Behavioral Procedures
        i. These procedures are largely adopted from the standards currently in use in Prof. Hulbert's Memory Dynamics Lab under Bard IRB Protocol 2015SEP18-HUL. The essential differences involve the addition of a new EEG (Emotiv) headset and real-time neurofeedback.
        ii. Tasks to be used in the current paradigm involve the presentation of words, images, or sounds via computer. Subjects will be asked to focus on, select, study, and/or make simple judgments about particular stimuli when prompted (e.g., "Where is the red 'L' on the screen?," "which of these images is different?," "how much do you remember this image on a scale from 1-5," or "is this picture of something that is alive?"). Responses will be spoken (into a microphone for offline or online coding of recorded responses or directed at the experimenter) or manual (e.g., button presses or mouse moves), allowing for the assessment of reaction time and/or accuracy measures. Participants may receive audio/visual indicators concerning their responses (e.g., a ding if they made the correct response). Between and after blocks (either immediately or after some delay that might be filled with unrelated puzzles, such as anagrams or math problems), participants' memory for certain stimuli presented during previous blocks may be tested through a recognition test (e.g., "did you see this item before?") a cued-recall test (e.g., "what was paired with this item?," or "which word did you see that started with this letter?") or an implicit test (e.g., the speed at which they are able to

name objects—some of which happen to have been presented earlier in the experiment—as they are gradually revealed on the screen or reading speed that is sensitive to whether participants recognize items from before). Participants will be told that they should respond as accurately as possible (and for some experiments, as quickly as possible). Detailed instructions and practice with the tasks will ensure that participants will not be confused about what to do throughout each phase of the experiment. This investigation will involve the use of innocuous (i.e., neither emotional, offensive, nor stressful) stimulus materials (see Appendix C.1 for example stimuli).

iii. To minimize fatigue, discomfort or eyestrain, subjects will be offered one or more rest periods during experiment sessions, which will last from approximately 10-120 minutes. During the rest periods, participants may stretch and/or close their eyes and rest for as long as they wish. Some experiments will consist of a single testing session while others will consist of multiple sessions that may take place on separate days (the relevant recruits would be informed of this prior to signing up for an appointment). To minimize discomfort, standardize the distance from the presented materials, and reduce unintended head movements during EEG recording, participants may be asked to use a table-mounted (nontoxic bakelite) chin rest (Cortech Solutions ET-EL-OP1KCR) placed in front of the computer screen. The chin rest will be wiped clean after each use.

At the end of the experiment, participants will be asked about their experience in the experiment (see Appendix D for example items from the post-experiment questionnaire). They will then be given a debriefing sheet that describes the hypothesis being tested and the logic of the experiment (a sample is provided in Appendix E). At this point, the experimenter will also answer any questions that the subject might have. Participants will be asked not to discuss the specifics of the experiment with other potential participants, so as to ensure that they would experience it in the same way.

b. EEG Procedures

i. In order to establish the EEG classifier that will be used to provide some participants with feedback based on these neural signals, participants will be asked to wear an EEG headset.

When participants arrive at the laboratory, they will be shown the EEG recording equipment and will receive a brief verbal summary of the experimental protocol. Then they will fill out a consent form that includes EEG recording procedures (see Appendix B).

EEG recordings primarily will be performed using a consumer-friendly, one-size-fits all, plastic headband (called EPOC+) made by the company Emotiv. To allow for validation testing, we have additionally included in this proposal relevant descriptions/procedures of another EEG system previously used in Bard Research (BioSemi).

Emotiv's (www.emotiv.com) EPOC+ neuroheadset is a product for the mass-market primarily designed to allow consumers to interact, via the headset, with one's personal computer. The technology has been used successfully to control computer games at home, in addition to its use in independent laboratory research (https://www.emotiv.com/category/independent-studies/). In addition to a gyroscope, accelerometer, and magnetometer to detect things like movement and orientation, the Emotiv headset features 14 electrode sensors. These sensors are to be used in conjunction with disposable felt pads to cover the electrodes, which can pick up the electrical brain activity that permeates the scalp. A new set of felt pads will be used for each participant and saturated with a mild saline solution (e.g., sterile contact lens solution) prior to each use. Pads will be disposed of after each participant has finished. Electrical activity from the brain, which permeates the scalp, can then be picked up by the electrodes and transmitted wirelessly over a restricted distance (using proprietary 2.4ghz wireless or Bluetooth Smart 4.0 LE), over radio frequencies to a yoked USB dongle, or sent via serial ports or wired USB on a nearby personal computer where it can be analyzed. The commercially available device, which runs on a small lithium battery, is in full FCC compliance—see appendix G—and has been approved by the International Electrotechnical Commission for Electrical Equipment's (IECEE) Certification Body for safety—see appendix H. Emotiv's wireless EEG EPOC+ is picture below.

ii. The EEG recording procedure for the BioSemi recording system described below is standard for electrophysiological laboratories in universities across the globe, including in Bard's Memory Dynamics Lab currently, and reflects the procedure recommended by the manufacturer of the electrophysiological supplies.

The BioSemi system uses electrodes that snap into place in mounts on a nylon cap that are fitted on the participant's head, along with up to 6 flat electrodes in plastic mounts, that are not embedded in the cap. The first step involves cleaning skin of excess skin oil over the areas in which the electrodes will be placed (behind the ears, on the forehead and cheek above/below one eye, and on the temples). Participants are invited to gently wash these areas using a paper towel, soap, and filtered water provided in the EEG preparation room. BioSemi's flat electrodes (sterilized after each use, see below) can then be attached to these locations with sterile, adhesive paper tabs. A drop of electrolyte gel (Signa Gel, a conductive saline solution that is hypoallergenic, bacteriostatic, non-gritty, and water soluble) is placed in the cup after the tab is attached. Only gel, a nylon cap, or the tab touch the participant's skin; the electrodes don't touch the skin directly. Each electrode is in contact only with the tab and the gel. The participants' electrical brain activity is picked up by the electrode through the gel.

The next step involves placing an electrode cap (containing 32 active electrodes) on the participant's head.

This cap or headband will fit over the head (caps are sized to fit the participant) and is held in place by a strap placed around the chin. Once the BioSemi cap is positioned on the head, a few drops of electrode gel are placed into the center of each electrode mount with a syringe that has a large, blunt plastic tip at the end of it. The end of the blunt syringe may then be used to gently work the gel into the hair and scalp right under the electrode to reduce the resistance of the contact to an acceptable level. At no point is the skin broken. There is no pain, although some subjects with very sensitive skin do occasionally report minor irritation. When this begins to occur, the experimenter uses less pressure. The blunt application devices and felt tips are sterile and disposable, each subject receiving fresh ones; the cap, headband, and electrodes are reused, but are washed and sterilized between uses according to the procedures described below. Once the gel is in place, a small electrode may then be fitted into the mounts built into the cap before applying it to the head. Again, no electrode ever touches the subject.

Each BioSemi electrode has an integrated amplifier built into it to reduce noise that could influence data quality. The cable protruding from the back of the electrode cap is then plugged into a low-power galvanically isolated anolog-to-digital post-amplifier box after the subject is positioned in the testing room (participants will be invited to touch a metal radiator, pipe, or electrically grounded mat to discharge/equalize any static charge they may be carrying). Moreover, numerous, redundant safety measures are built into the device. The BioSemi ActiveTwo system has a “Driven Right Leg” (DRL) circuit with current limiter. Besides reducing the Common Mode voltage, the DRL also protects the subject for defects in the amplifier. If one of the input stages in the active electrodes would break down, and the electrode input would become shorted to one of the active electrode supply rails (0V or 4V), the current limiting resistor in the DRL protects the subject. In the very unlikely case that two active electrodes would fail simultaneously and that one electrode would be connected to the 0V and the other electrode would be connected to the 4V then dangerous current would be possible in spite of the DRL current limiter. Therefore, BioSemi has integrated an extra protection in the ActiveTwo analog-to-digital box, which only enables the powers when no errors are detected.

The subject is protected for leakage currents from the mains supply by the isolation barrier between the amplifier and the PC: the optical fiber data-link combined with battery power supply provides complete safety. Leakage currents are well below the measurement accuracy (< 1 uA). In case of the optional mains supply, the low capacitance of the used DC-DC converters limits the currents to less than 10 uA.

The DRL provides an additional safeguard for mains supply currents when someone would by mistake make a ground (earth) connection to the amplifier. Note that this would be very difficult, since the "saboteur" would have to open the cabinet to do this since all conductors on the outside of the cabinet are either electrode inputs or protected shield outputs.

On the standard DRL circuit, the output current is limited by a 500 kOhm resistor inside the DRL integrator loop. The DRL integrator runs on a 4V supply. This results in a maximum error current of 10 uA (highly unlike worst case scenario: the DRL integrator swings to 4V and one input shorts to 0V, or vice versa).

Finally, the cabinet is constructed in such a way that the subject can never touch unprotected low-impedance points, such as ground planes, power supply rails or amplifier outputs. The analog-to-digital box passes brain signal on to a computer in the adjacent control room through an optical cable, which records the data. Once the electrode offsets are brought down to an acceptable level, the experiment can begin.

While EEG signals are being recorded from participants, they will be asked to perform tasks following the aforementioned descriptions under *Behavioral Procedures*. At certain critical points during each trial, the participant may be asked to refrain from blinking or moving their eyes for periods lasting from 1-5 seconds, as eye blinks and movements create electrical noise which contaminate the recorded EEG.

As space is limited and the presence of the experimenter in the testing room may add data noise in the room, participants may be monitored remotely via a video intercom system. This system will allow for two-way audio communication (to deliver instructions and should the participant need to get in touch with the experimenter in the control room for any reason). Participants will be informed prior to the experiment that the audio/video monitoring is being done to ensure that they are comfortable and on task throughout the experiment, will only be seen by approved/trained research personnel, and will NOT be recorded.

The BioSemi preparation ("EEG capping") of the participant typically takes 20-30 minutes, the Emotiv EPOC+ requires 5-15 minutes for fitting and connection procedures. Subsequent data collection requires 10-120 minutes. Short breaks are provided during the data collection phase.

When data collection is complete, the BioSemi cap and electrodes are removed from the participant's head and the electrode gel is dabbed off with tissues. Sometimes, a little gel remains in the hair, but the participant is instructed that this will rinse out easily with their next shower. Likewise, The EPOC+ may leave small amounts of residual saline solution which similarly washes away with water and does not irritate the scalp. The experimenter can offer the participant towels and shampoo to be used in the available salon-style sink with head basin and sprayer. Then the participant completes the post-experiment questionnaire, is debriefed, and dismissed.

At this point, the BioSemi cap and electrodes would be washed and sterilized. The cap is submerged in a bucket of warm water and Ivory detergent to clean it for 10 minutes. Any remaining electrode gel is removed with a water sprayer. The electrodes are rinsed with warm water and the sprayer to remove any electrode gel adhering to them. Then, the cap is sprayed with a standard hydrogen peroxide disinfectant spray in order to sterilize them. Everything is left to dry before subsequent reuse. Similarly, the Emotiv EPOC+ headband is wiped down and the felt tips discarded after the participant has completed their involvement in the study. A set of felt tips will be used for each participant, but this set, after use, may be stored in the event of requiring further sessions with that participant. If so, the felt tips will be sealed in plastic bags, labeled with the participant number and date, and put in a locked cabinet.

iii. A multivariate classifier (a software algorithm trained on EEG data from the current or prior experiment sessions) may be used to determine the number, type, or ordering of the trials or to predict future behavioral outcomes. This classifier "learns" to identify when the current brain state matches one or more previously established brain states (e.g., one associated with paying attention), according to previously recorded data. For instance, the algorithm might wait to present a stimulus or change a stimulus until the participant is detected to be in a particular brain state (e.g., "high-learning state"). Later, the memorability for that stimulus could be compared to that of a stimulus delivered when the same participant was in a different brain state (e.g., "low-learning state"), or the stimulus could be withheld or altered until a particular state is recognized (see Appendix C.2). Such information could be used to help guide learners into adopting brain states more/less conducive to learning, as well as to assess the effectiveness of various learning regimens and classification algorithms.

10. **Approximately how many individuals do you expect to participate in your study?** Depending on counterbalancing factors, the level of noise, and statistical power, I expect I will need between 6-12 valid participants in each condition of the experiment in question for preliminary testing. Data collection will be ongoing throughout the year and may continue through future renewals of this protocol, subject to IRB review.

11. **Please describe any risks and benefits your research may have for your participants. (For example, one study's risks might include minor emotional discomfort and eyestrain. The same study's benefits might include satisfaction from contributing to scientific knowledge and greater self-awareness.)**
    a. This protocol presents minimal risk for participants. We make every effort to reduce the possible fatigue that may arise from performing a cognitive task for the duration of the session by including regular breaks and resting equipment. We also make every effort to ensure that no discomfort occurs as a results of

EEG procedure or equipment and regularly ask for participants to indicate their level of comfort.

b. For any procedure requiring the placement of electrodes on the skin or scalp, there's an extremely minor risk of very slight skin irritation in a very small number of participants. This requires no treatment and disappears within several minutes after the application of the electrode gel is complete. Participants' skin is never broken. Electrodes will be applied (and later removed) with the utmost care and attention to the participants' comfort level from moment to moment. All the reused materials that touch the subject's skin are thoroughly cleaned between uses, according to standard electrophysiological lab procedures. There is no risk of electrical shock. Details can be found in *EEG Procedures*, above.

c. Sometimes, participants may ask if there are any abnormalities in their EEGs. Those participants who ask this question will be told that this study (and indeed the hardware used to collect their brain data) is not designed for clinical diagnosis and that such diagnoses can only be made by a neurologist or clinical electrophysiologist. Thus, they will be informed that there are not any diagnostic conclusions we can draw from these data.

d. While there are no direct benefits to participants, compensated participants may indirectly benefit from learning about the research process (especially true for Bard psychology students), as well as about the background motivating the present work. Specifically, their experience and the provided debriefing information may help them identify strategies that benefit their ability to flexibly control attention and memory systems to better meet their learning goals. Moreover, it is hoped that participants will experience satisfaction for having contributed to the growing scientific body of knowledge emanating from Bard. On a societal level, the present research promises to help the scientific community—at Bard and beyond—understand the basic mechanisms of memory and attention. To the extent that we understand such basic cognitive processes, we are in a better position to design new instructional and learning technologies and methodologies to foster learning in both healthy and learning-impaired populations.

12. **Have you prepared a consent form and emailed it as an attachment to IRB@bard.edu?** Yes, the consent form has been emailed to the IRB, along with the rest of the supplements.

13. **Please include here the verbal description of the consent process (how you will explain the consent form and the consent process to your participants):**

a. Recruits will initially be told that the study is investigating how electrophysiological signals correspond to their ability to perform computerized tasks. They'll be informed that the experimenter will provide them with all the necessary instructions and walk them through each step of the experiment, as well as a full debriefing after the experiment is over. After confirming that they are eligible for the experiment, the experimenter will then provide a brief oral description of the tasks they'll be asked to perform and equipment to be used during the experiment. They will be shown the equipment and given a description of how it will be used in the experiment to make sure that they are comfortable with the equipment and procedures. Should they indicate their willingness to participate, all participants will be provided a written informed consent agreement that describes the study in more detail. They will then be asked to repeat back, in their own words, the procedure laid out in the consent form and to verbally answer a set of basic questions establishing their understanding and their right to withdraw from the study at any point without penalty. Provided all parties reach a common understanding, the participant will be invited to sign the consent agreement. All participants will be told that they are welcome to ask questions about the experiment both before and after the experimental session and pointed to the additional contact information provided on the consent/debriefing forms.

14. **If your project will require that you use only a verbal consent process (no written consent forms), please describe why this process is necessary, how verbal consent will be obtained, and any additional precautions you will take to ensure the confidentiality of your participants.** Not applicable
15. **What procedures will you use to ensure that the information your participants provide will remain confidential?** Email addresses (collected to contact and schedule participant sessions and to enter themselves into the raffle) will be kept separately from the behavioral and electrophysiological data collected over the course of the experiment. They will be linked to the rest of the participant data by an arbitrary string of numbers (i.e., a participant number), with the linking document stored separately on a password-protected computer maintained by the trained and certified research team in order to maintain confidentiality. Individually identifiable data will not be released to anyone outside the research team without the written consent of the participant. If any information obtained from this study is published, the article will be written so that the identity of all subjects will remain confidential. Any audio files with participant responses will similarly be stored in a secure manner within the

confines of the laboratory. Signed consent forms will be stored separately from the data, in a locked filing cabinet accessible only to members of the research team who are certified to work with human subjects (defined above in question #8). All study materials will be coded and entered into password-protected computer files. Any publication or conference presentation stemming from the research in question would avoid the inclusion of any identifying participant information.

16. **Will it be necessary to use deception with your participants at any time during this research? Please note: withholding details about the specifics of one's hypothesis does not constitute deception. However, misleading participants about the nature of the research question or about the nature of the task they will be completing does constitute deception.** Yes
17. **If your project study includes deception, please describe here the process you will use, why the deception is necessary, and a full description of your debriefing procedures.**
    a. All recruits will be told that the study is designed to investigate how electrophysiological signals correspond to their ability to perform computerized tasks. While this is true, certain additional information may be withheld from participants in order to test questions about effective learning strategies. In particular, participants may not be told at the outset that their memory will be tested as part of the experiment. Moreover, the experimenters may implicitly or explicitly indicate that there will *not* be a memory test. Many of the forms of learning and memory to be investigated are *incidental*, such that participants learn without trying or even being aware that learning is happening. This aspect of the research is critical, since explicitly trying to learn/memorize is thought to draw from partially distinct cognitive (and biological) resources (Rugg et al., 1998). In fact, past research has shown that trying to learn interferes with incidental forms of learning (Roediger, 1990). Therefore, telling participants up front that they will be tested would invalidate some of the hypotheses being investigated. When possible, participants will be given partial information that there will be a subsequent part of the experiment involving a different task and that they'll be given new instructions at that point. Regardless, participants will be fully debriefed about the stages of the experiment, the full hypotheses being tested, and how the different tasks help address these hypotheses (see below). Furthermore, participants who are given surprise memory tests will be given the opportunity during the debriefing session to withdraw their consent. Should they wish it, we will

discard their data as requested. This minor withholding of information does not expose participants to any additional risks.

b. Experiments may involve withholding additional information or providing misinformation about the use of EEG when necessary to test the effectiveness of neurofeedback. Some participants may be given partial information about the nature of the neurofeedback they will receive in the tasks. For example, in the phase of the research that involves the provision of feedback about brain activity, some participants may receive feedback that is distorted to appear stronger or weaker than the actual activity, while other participants may receive feedback about activity from another subject's brain (without any way to identify from whom those brainwaves were generated), and still other participants may receive feedback drawn from control regions of the scalp or features unrelated to attention or memory. We have no reason to anticipate that this type of deception should involve any additional risks. These procedures are standard for testing neurofeedback, including in the undergraduate thesis work at Princeton University, which was supervised by my current faculty adviser, Justin Hulbert. The deception is absolutely critical to these experiments, as we are aiming to test how accurate feedback alters learning; if we gave all participants diagnostic feedback, we could not determine the necessity of the feedback or its critical components (e.g., perhaps any type of practice with the task would lead to improvements). Explaining the nature of the feedback to participants up front would invalidate some of the hypotheses being investigated, since to be an effective control, participants must believe that the feedback is diagnostic. Otherwise, they may be tempted to change their strategy or disregard the feedback entirely. All participants in these conditions will be made aware of the deception after the conclusion of their participation in the research and will be offered the opportunity to retract their data upon reveal of this deception (see below).

After completing their involvement in the research, all participants will be asked some general questions (e.g., "What do you think this experiment was testing?," "Did you use any particular strategy to accomplish your task(s)?," "How do you feel you did on the task?," "Do you have any other comments about this experiment?"). These questions will help assess whether the experiment[er] met their expectations, whether the instructions had been sufficiently clear, and that they had a positive experience. They will then be given a debriefing sheet that describes, in detail, the full set of hypotheses being addressed, how the experiment addresses these hypotheses, the broader significance of the research, and how to get in touch with relevant parties should they have any further questions or concerns. The experimenter will answer any questions that the participant might have. A sample debriefing sheet is attached as Appendix E. For participants who were given a surprise memory test or neurofeedback, the debriefing will include the following statement: "This experiment required us to withhold information from you in order to avoid contaminating the results. In particular, we did not tell you in advance about the surprise memory test. Intentionally trying to learn is a very different process than the learning that incidentally occurs when you perform a task. In fact, past research has shown that trying to learn can interfere with more incidental forms of learning. Furthermore, the neurofeedback you observed may not have been related to the purported cognitive processes expressed by your experimenter. This would have been done in order to establish that real neurofeedback has benefits above and beyond that of feedback unrelated to your attentional/memory brain states, and we require some participants to act as a control in order to establish whether our hypotheses are correct. Therefore, telling you up front that you would be tested on these materials and that you may receive sham neurofeedback could invalidate the hypotheses being investigated. We apologize for withholding this information about the experiment before you participated. Please let your researcher know if we may still use your data in our study."

c. If the participant indicates that they do not want their data used in our research, we will discard their data. Regardless, all participants will be thanked and will be compensated according to the format established during the intake process. Participants will also be asked not to discuss the specifics of the experiment with other potential participants, so as to ensure that they would experience it in the same way.

18. **For projects not using deception, please include your debriefing statement. (This is information you provide to the participant at the end of your study to explain your research question more fully than you may have been able to do at the beginning of the study.) All studies must include a debriefing statement. Be sure to give participants the opportunity to ask any additional questions they may have about the study.** See Appendix E for a sample debriefing statement.

SECTION 3

1. **If you will be conducting interviews in a language other than English, will you conduct all of the interviews yourself, or will you have the assistance of a translator?** Not applicable.
2. **If you will be using the assistance of a translator, that individual must also certify that he or she is familiar with human subject protocol and has completed the online training course. Please respond whether you have found an IRB-certified translator.** Not applicable.
3. **If you have not yet found a translator, do you agree that when you do find a translator, you will make sure that person will also agree to use standard protocol for the treatment of human subjects, and that the individual's training certificate will be submitted to the IRB records before you begin collecting data?** Not applicable.
4. **If your recruitment materials or consent forms will be presented in languages other than English, please translate these documents and email copies at attachments to IRB@bard.edu.** Not applicable.
5. **I have submitted all my translated materials.** Not applicable.
6. **I have submitted a copy of my video consent form.** Not applicable.

SECTION 4

1. **If you are a graduate or undergraduate student, has your adviser seen and approved your application?** Yes.
    a. **If you have not already done so, you must ask your adviser to email a statement on your behalf to IRB@bard.edu The statement should read, "I have reviewed [your name]'s proposal and I will oversee this research in its entirety."**
2. **Please read the following statement carefully: "I have read the Bard IRB policy on the treatment of human research participants. I will comply with the informed consent requirement, and I will inform the IRB if significant changes are made in the proposed study. I certify that all of the information contained in this proposal is truthful." Submitting this form means that you affirm the statement above and will comply with the content. This counts as your legally binding signature.**

I concur with the above,

Noah Libby

Appendices

Appendix A: Sample recruitment text

Thanks so much for your interest in this study! This investigation will observe your ability to put your brain to the test as you perform various computer-based tasks. We're looking for healthy adult participants who would be willing to wear a portable headset device that can read their brainwaves (called electroencephalograms or EEG), so that we can better determine how your brain allows you to perform the tasks. In some cases, these brainwaves can even be used to control some stuff on the computer!

To be eligible, you MUST:

- Be between 18-35 years old
- Have normal or corrected-to-normal color vision (glasses and contacts are OK)
- Be a native English speaker
- NOT have hairstyles that will hinder brainwave recording (such as dreadlocks)
- NOT have a diagnosed attention deficit, learning disability, or neurological condition
- Be willing and able to sit still and keep your eyes/attention focused and maintain fixation for extended periods during the duration of the experiment, without discomfort or stress
- NOT have participated in previous **attention** experiments at Bard (last academic year, over the summer, or during the fall 2016 semester)

If you meet ALL of the above eligibility criteria and remain interested in participating, you can browse the available appointment times and book one on the following page: ____________

Be sure to choose an appointment time that allows you to arrive at the experiment rested and ready to go. It's also a good idea to give yourself 15 extra minutes before you need to be somewhere else, in case the experiment runs slightly over its scheduled time.

If you don't find a posted time that works for you, more slots will be posted in the coming days. Check back on the website soon!

If you have any questions, please email nl8800@bard.edu with "Experiment Question" in the subject line.

Thanks again for your interest!

Appendix B: Consent form

INFORMED CONSENT AGREEMENT

Protocol number: **Expires:**
Study title: Real-Time Neurofeedback

Principal investigator: Noah Libby

You are being asked to take part in a research experiment at Bard College that seeks to learn about how different brain states are associated with performance abilities on certain computer-based tasks.

To decide whether or not you wish to participate, you should know enough about its risks and benefits to make an informed judgment. This consent form gives you information about the research study, and the experimenter will provide you with additional information about the specific tasks that you will be performing. Once you are ready, you will be asked if you wish to participate and, if so, you will sign the consent form. You can choose not to participate, and you can choose to end your participation at any time during the study.

What you will do in this study: Should you be eligible and decide to participate, you will be asked to make simple judgments about written (words), visual (images), or auditory materials (sounds) presented by a computer by pressing buttons, moving a mouse, or speaking out loud into a microphone that will capture your responses. The researcher will offer detailed instructions to guide you through each part of the experiment and answer any questions you may have about the procedure. After the experiment, you will then be asked to fill in a brief questionnaire about the experiment and given an opportunity to ask any remaining questions that you may have.

During this task, we may record the tiny electrical signals generated by your brain (so-called brainwaves). To do this, small, sterilized electrodes (or ones buffered by clean, disposable felt pads) will be placed over your head using a small amount of gel or saline solution that helps transfer the signal from your body to the recording electrodes, with no risk that they could shock you. The whole process is non-invasive and not painful. You are encouraged to keep the researcher informed of your continued comfort during the application of, removal of, and recording using these measurement devices. These data may be used to provide you with feedback about your brain state and may also alter stimuli presented to you on the screen.

It is expected that the first 5-15 minutes of the experiment will be spent preparing you and the measurement devices, leaving the rest of session for the actual task and cleanup. The total time for a session is not expected to run longer than 2 hours. You will be offered the opportunity to take breaks throughout. You may be invited back for additional sessions, but similarly, you can end participation at any time or opt out of future sessions/contacts without penalty. Should you ever decide to end your participation early, you are encouraged to simply let the experimenter know. All the information and responses collected during the experiment will be deleted upon request.

Risks and benefits: There are no health risks associated with this study and most participants report having a positive experience. Experiment sessions are kept as short as possible, and every attempt is made to ensure that participants are kept as comfortable as possible throughout. Participants are reminded that, should they become fatigued or in any way uncomfortable during the experiment, they may ask for a break or withdraw at any time without penalty.

After the experiment, participants may prefer, for appearance reasons, to wash off remnants of the completely harmless electrode gel or solution with the provided soap and water.

The words, images, and sounds participants may encounter during the experiment are intended to be neutral, non-threatening, and inoffensive. If you are a student at Bard College and find that any aspect of the experiment caused you distress, you are encouraged to contact the Bard Counseling Center at 845-758-7433 during normal business hours or at 845-758-7777 after hours or on weekends. Even if you are not a Bard College student but find yourself experiencing significant distress, please contact the National Alliance on Mental Illness (NAMI) at 1-800-950-NAMI (6264).

While this research experiment may not provide participants with any direct benefits, the data collected from this study may help improve the scientific understanding of how to effectively control the focus of attention and the results of doing so. Additionally, we hope that some participants may come away from this experiment with a better grasp of how signals from the brain can influence your everyday life. Moreover, the researchers hope that participants gain insight into the research process at Bard College and beyond through their involvement with this work.

The experimenter will tell you more about the study and our hypotheses at the end of the session.

Compensation: In exchange for participating in this experiment, you may be offered a token piece of candy and raffle entries for Amazon gift cards (with the pot ranging from $25-50), with winners selected at random by May 22, 2019, plus any bonuses introduced during the procedure. Should you be invited back for additional experiment sessions, you will be compensated

$5/hour for every hour in these additional sessions.

Your rights as a participant: Your participation in this experiment is completely voluntary, and you may withdraw from the experiment at any time without penalty. You will still receive any stated compensation for your participation up until that point. You may withdraw by informing the experimenter that you no longer wish to participate.

Confidentiality: All records from this study will be kept confidential. Your responses will be assigned an arbitrary participant number and kept strictly private, shared only with the investigator and trained members of the research team (faculty members and undergraduates at Bard College) who have been certified for work with human participants. We will not include any information that will make it possible to identify you in any report we might publish, including the resulting Senior Project, which will be publicly accessible at Bard College's Stevenson Library and on the online thesis repository, the Digital Commons. Research records will be stored securely in a locked cabinet and/or on password-protected computers.

If you have questions about this study, please ask your researcher, Noah Libby (nl8800@bard.edu), or contact Dr. Justin Hulbert (Psychology Program, Bard College, Annandale-on-Hudson, NY 12504; jhulbert@bard.edu). If you have questions about your rights as a research participant, please contact the Bard College Institutional Review Board at irb@bard.edu.

STATEMENT OF CONSENT:

"The purpose of this study, procedures to be followed, and the risks and benefits have been explained to me. I have been given an opportunity to ask questions, and my questions have been answered to my satisfaction. I have been told whom to contact if I have additional questions. I have read this consent form and agree to be in this study, with the understanding that I may withdraw at any time."

By signing below, I agree with the above **statement of consent** and further certify that I am at least 18 years of age.

______________________________ __________

Participant signature Date

____

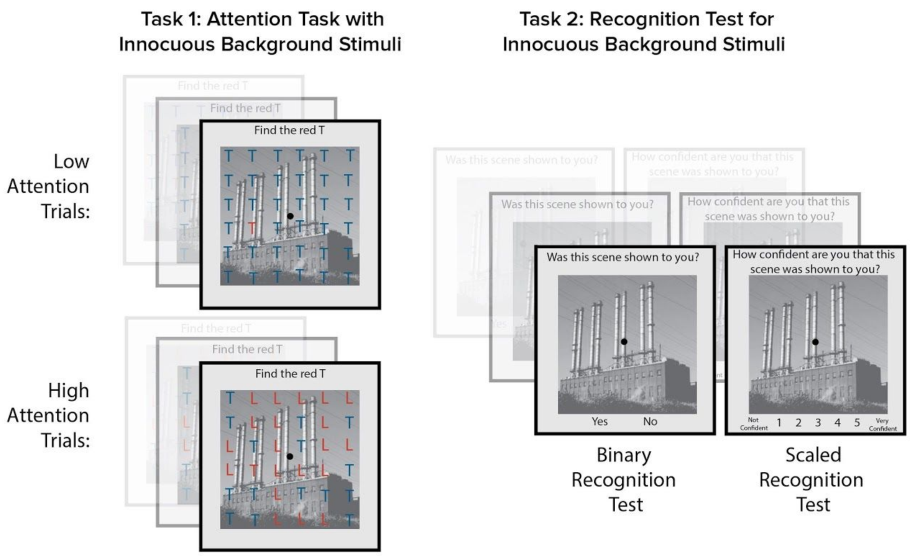
Participant name (printed)

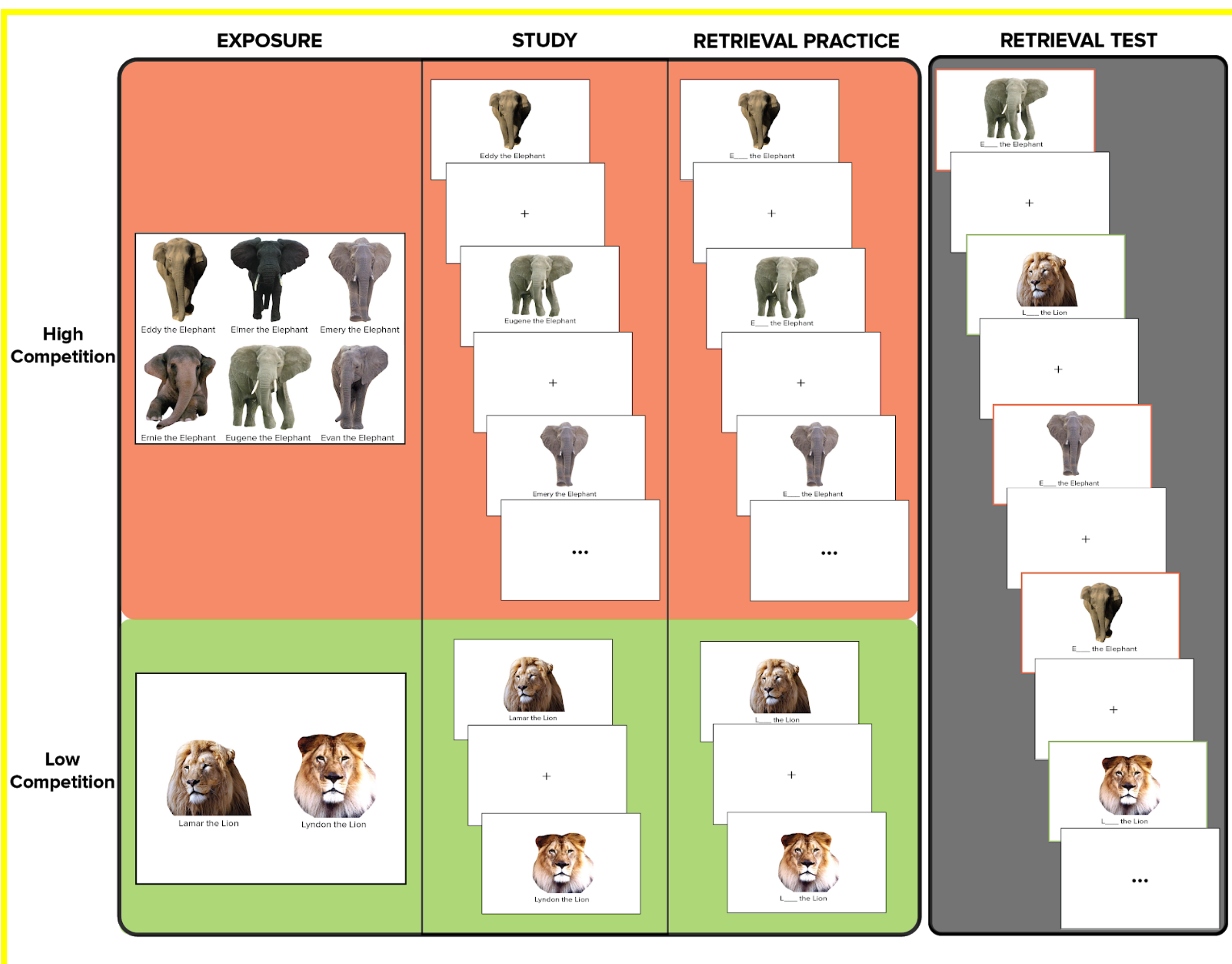
____

Experimenter signature

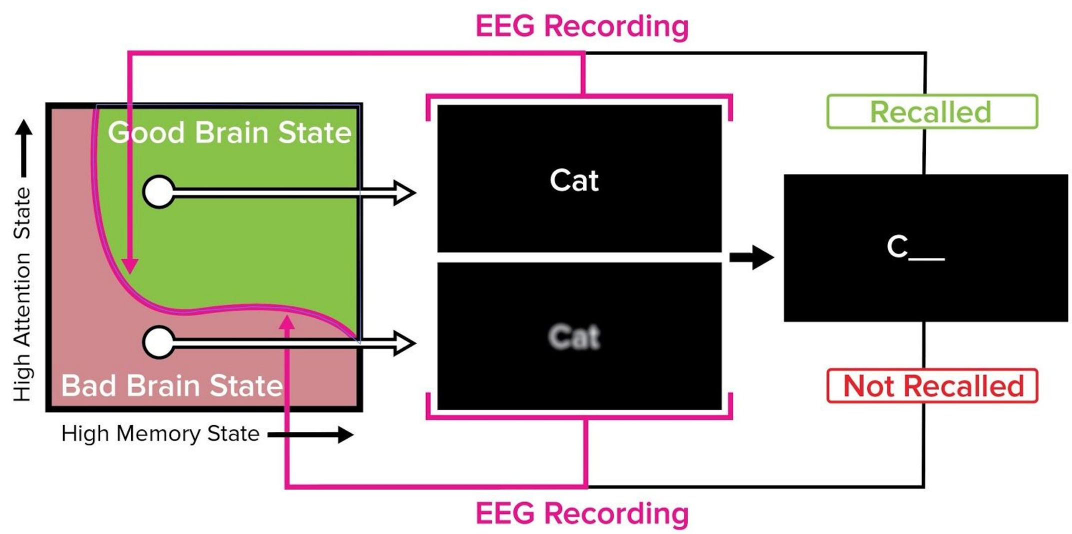
Appendix C: Example stimuli & methods overview

1. Classifier "Training" Procedures





2. Neurofeedback Procedure



Appendix D: Sample items from post-experiment questionnaire

Task 1:

• How often did you pay attention to the images during the study period?

Never	Rarely	Sometimes	Often	Always
0	1	2	3	4

How often did you pay attention to the background "distractor" images while completing the visual search tasks?

Never	Rarely	Sometimes	Often	Always
0	1	2	3	4

• Did you ever pay attention to the background "distractor" images on purpose while completing the visual search tasks?

Never	Rarely	Sometimes	Often	Always
0	1	2	3	4

• To what extent did you expect to be tested for the "distractor" images before/while completing the visual search task?

Not at all	A little	A bit	Quite a bit	A lot
0	1	2	3	4

Task 2:

• Do you think the recognition test captured your memory for the images?

Not at all	A little	A bit	A lot	Completely
0	1	2	3	4

Task 3:

• How much do you think the neurofeedback affected your learning?

Not at all	A little	A bit	Quite a bit	A lot
0	1	2	3	4

• Do you believe that the neurofeedback was relevant to your learning?

Not at all	A little	A bit	Quite a bit	A lot
0	1	2	3	4

• Do you believe that the neurofeedback reflected your cognitive states?

Not at all	A little	A bit	Quite a bit	A lot
0	1	2	3	4

• Did you identify or use strategies to change the neurofeedback?

Not at all	A little	A bit	Quite a bit	A lot
0	1	2	3	4

• What strategies did you find most useful?

• Do you think that these strategies are helpful for learning?

Not at all	A little	A bit	Quite a bit	A lot
0	1	2	3	4

•Are there strategies that you found less useful?

• How accurate were your predictions for how many items you would remember?

Very inaccurate	Somewhat inaccurate	Somewhat accurate	Very accurate
1	2	3	4

General:

• Last night, how many hours of sleep did you get? (estimate)

$\leq$2 3 4 5 6 7 8 9 10 $\geq$11

• How often did you experience stress completing this experiment?

Never	Rarely	Sometimes	Often	Always
0	1	2	3	4

• Do you feel like what you experienced might help you learn in the future?

NO 0 1 2 3 4 YES

• Do you have other comments or questions?

To be administered after debriefing:

• Now that you know the neurofeedback might not have reflected your real, current brain state, do you believe that the neurofeedback you received was accurate?

Not at all	A little	A bit	Quite a bit	A lot
0	1	2	3	4

• To the extent that you believed that your feedback did not reflect your real, current brain state, was there anything particular in the experience that led you to suspect this?

Appendix E: Sample debriefing form

Study title: Paying Attention to Real-Time Neurofeedback

Principal investigator: Noah Libby (nl8800@bard.edu)

Thank you for participating in this experiment. This research is designed to explore the basic mechanisms underlying attentional control and memory. By conducting this study, we hope to learn more about how people might learn to better focus their cognitive state in a way that allows them to remember what they want to remember and when they want to remember.

In the first part of the experiment, we asked you to focus your attention on one or more primary tasks. Although we were interested in how well you performed on the primary task(s) by controlling your attention and the electrophysiological data recorded during this time, we were also interested in your performance and the electrophysiological data associated with memory for stimuli that appeared between, in, or around this attentional task, and how this information could be used to predict later memory. To examine this, we may have introduced "distractor" materials and later surprised you with a memory test for these distractors.

The reason for withholding information about the upcoming memory task was that we required a measure of memory for events that took place without the intention for these events to be remembered. By combining the electrophysiological data associated with the attentional manipulation in the first task and the "incidental memory" correlates attained by relating the stimuli you remembered and forgot with the related electrophysiological data, we hoped to present you with neurofeedback that varied between being controlled by your personalized "high attention" brain state correlates and your "high incidental memory" brain state correlates. Because retention of information requires both attentional and memory processes, we hypothesize that ideal neurofeedback for explicit memory would be presented based on some combination of your electrophysiological correlates of these two tasks.

By researching the nature of these combined brain states, as well as how individual memories may compete with one another as they are being learned, and utilizing them for feedback, we hope to increase our ability to control our retention for material. For example, students might be able to use this device and computer algorithm studying for an exam.

This experiment required us to withhold information from you in order to avoid contaminating the results. In particular, we did not tell you in advance about the surprise memory test. Intentionally trying to learn is a very different process than the learning that incidentally occurs when you perform a task. In fact, past research has shown that trying to learn can interfere with more incidental forms of learning. Furthermore, the neurofeedback you may have *observed may not have been related to the purported cognitive processes expressed by your experimenter. This would have been done in order to establish that real neurofeedback has benefits above and beyond that of feedback unrelated to your attentional/memory brain states, and we require some participants to act as a control in order to establish whether our hypotheses are correct. Therefore, telling you up front that you would be tested on these materials and that you may receive sham neurofeedback could invalidate the hypotheses being investigated. We apologize for withholding this information about the experiment before you participated. Please let your researcher know if we may still use your data in our study.*

Regardless, if you have any questions or concerns, you may ask your experimenter, Noah Libby in person or at nl8800@bard.edu, or feel free to contact his faculty supervisor, Dr. Justin C. Hulbert, at jhulbert@bard.edu. You may email the Bard College Institutional Review Board at irb@bard.edu for questions about your rights as a participant.

Again, we thank you for your participation. If you know of any friends or acquaintances that are eligible to participate in this study, we kindly request that you not discuss it with them until after they have had the opportunity to participate. Prior knowledge of questions asked during the study can invalidate the results. We greatly appreciate your cooperation.

Bard Institutional Review Board	Bard Counseling Center	National Alliance on Mental Illness Hotline
irb@bard.edu	845-758-7433 / 7777	1-800-950-NAMI (6264)

Appendix F: NIH human participant protection education certificate

Certificate of Completion

The National Institutes of Health (NIH) Office of Extramural Research certifies that **Zall Hirschstein** successfully completed the NIH Web-based training course "Protecting Human Research Participants".

Date of completion: 02/14/2016.

Certification Number: 2003376.

Certificate of Completion

The National Institutes of Health (NIH) Office of Extramural Research certifies that **Justin Hulbert** successfully completed the NIH Web-based training course “Protecting Human Research Participants”.

Date of completion: 08/12/2015

Certification Number: 1812737

Certificate of Completion

The National Institutes of Health (NIH) Office of Extramural Research certifies that **Noah Libby** successfully completed the NIH Web-based training course "Protecting Human Research Participants".

Date of completion: 11/02/2016.

Certification Number: 2223569.

Appendix G: Emotiv declaration of conformity

DECLARATION OF CONFORMITY

according to ISO/IEC 17050-1 and EN 17050-1

DoC #: EPOC_USB_01
November 8th 2011

Supplier's Name: Emotiv Limited
Supplier's Address:
Room 611 Fook Cheong Building
63 Hoi Yuen Road
Kwun Tong, Hong Kong

declares that the product

Product Name and Model: Emotiv EPOC Model 1.0
Product description: EPOC Neuroheadset, USB-01 Transceiver, Hydrator Pack + charger or charge cable

conforms to the following Product Specifications and Regulations:

EMC and Telecom: Class B
ETSI EN 300 440 2 V1.4.1
EN 301 489-1
EN 301 489-3
AS/NZS CISPR22 :2009
AS/NZS 4268 :2008
FCC CFR 47 Part 15C (identifiers XUE-EPOC01, XUE-USBD01)

Safety:
EN 60950 1:2006
IEC 60950-1:2005 (2nd Edition)
AS/NZS 60950.1:2003 including amendments 1, 2 & 3
CB Certificate JPTUV-029914 (TUV Rheinland)

The product herewith complies with the requirements of the Low Voltage Directive 2006/95/EC, the EMC Directive 2004/108/EC, the R&TTE Directive 1999/5/EC, and carries the CE mark accordingly.

Technical File:
The technical file is maintained at
Emotiv Limited
Suite 145, NIC Building
4 Cornwallis St
Eveleigh NSW 2015
AUSTRALIA

Signed for and on behalf of Emotiv on November 8th, 2011.

Geoffrey Mackellar
Chief Technology Officer, Emotiv Limited

Appendix H: Emotiv IECEE approval

IEC IECEE CB SCHEME

Ref. Certif. No.
JPTUV-029914

IEC SYSTEM FOR MUTUAL RECOGNITION OF TEST CERTIFICATES FOR ELECTRICAL EQUIPMENT (IECEE) CB SCHEME

SYSTEME CEI D'ACCEPTATION MUTUELLE DE CERTIFICATS D'ESSAIS DES EQUIPEMENTS ELECTRIQUES (IECEE) METHODE OC

CB TEST CERTIFICATE
CERTIFICAT D'ESSAI OC

Product Produit	Consumer EEG wireless headset for brain computer interface
Name and address of the applicant Nom et adresse du demandeur	Emotiv Systems Inc. Suite 217/10 Nic Building, Australian Technology Park 4, Cornwawsst, Eveleigh 2015, Australia
Name and address of the manufacturer Nom et adresse du fabricant	Dae Kyung Philippines Co., Ltd Lot 1-6 Block 20 Phase IV main Avenue, EPZA Rosario Cavite 4106, Philippines
Name and address of the factory Nom et adresse de l'usine	Dae Kyung Philippines Co., Ltd Lot 1-6 Block 20 Phase IV main Avenue, EPZA Rosario Cavite 4106, Philippines
Rating and principal characteristics Valeurs nominales et caractéristiques principales	Powered via USB connection Class III
Trade mark (if any) Marque de fabrique (si elle existe)	Emotiv EPOC™
Model/type Ref. Ref. de type	1.0
Additional information (if necessary) Information complémentaire (si nécessaire)	
A sample of the product was tested and found to be in conformity with Un échantillon de ce produit a été essayé et a été considéré conforme à la	IEC 60950-1:2005
As shown in the Test Report Ref. No. which forms part of this Certificate Comme indiqué dans le Rapport d'essais numéro de référence qui constitue une partie de ce Certificat	12020403 001

This CB Test Certificate is issued by the National Certification Body
Ce Certificat d'essai OC est établi par l'Organisme National de Certification

TÜVRheinland®

TÜV Rheinland Japan Ltd.
Global Technology Assessment Center
4-25-2 Kita-Yamata, Tsuzuki-ku
Yokohama 224-0021 Japan
Phone + 81 45 914-3888
Fax + 81 45 914-3354
Mail: info@jpn.tuv.com
Web: www.tuv.com

Date: 10.12.2009

Signature: Dipl.-Ing. S. Hartter

10/061 CB 1.07

Appendix I: References

DeBettencourt, M. T., Cohen, J. D., Lee, R. F., Norman, K. A., & Turk-Browne, N. B. (2015). Closed-loop training of attention with real-time brain imaging. *Nature Neuroscience*, *18*(3), 470–478. https://doi.org/10.1038/nn.3940

Kober, S. E., Schweiger, D., Witte, M., Reichert, J. L., Grieshofer, P., Neuper, C., & Wood, G. (2015). Specific effects of EEG based neurofeedback training on memory functions in post-stroke victims. *Journal of NeuroEngineering and Rehabilitation*, *12*(1), 107. https://doi.org/10.1186/s12984-015-0105-6

Roediger, H. L. (1990). Implicit memory. Retention without remembering. *The American Psychologist*, *45*(9), 1043–1056. https://doi.org/10.1037/0003-066X.45.9.1043

Rugg, M. D., Mark, R. E., Walla, P., Schloerscheidt, A. M., Birch, C. S., & Allan, K. (1998). Dissociation of the neural correlates of implicit and explicit memory. *Nature*, *392*(6676), 595–598. https://doi.org/10.1038/33396