Using Computational Modeling to Assess Use of Cognitive Strategies

Michael J. Haass and Laura E. Matzen

Sandia National Laboratories,
P.O. Box 5800, MS 1188,
Albuquerque, NM, USA 87185
{mjhaass,lematze}@sandia.gov

Abstract. Although there are many strategies and techniques that can improve memory, cognitive biases generally lead people to choose suboptimal memory strategies. In this study, participants were asked to memorize words while their brain activity was recorded using electroencephalography (EEG). The participants' memory performance and EEG data revealed that a self-testing (retrieval practice) strategy could improve memory. The majority of the participants did not use self-testing, but computational modeling revealed that a subset of the participants had brain activity that was consistent with this optimal strategy. We developed a model that characterized the brain activity associated with passive study and with explicit memory testing. We used that model to predict which participants adopted a self-testing strategy, and then evaluated the behavioral performance of those participants. This analysis revealed that, as predicted, the participants whose brain activity was consistent with a self-testing strategy had better memory performance at test.

Keywords: Memory, computational modeling, electroencephalography.

1 Introduction

Memory underlies and supports all forms of high-level cognition and accurate memory is essential to good decision making. However, human memory is extremely fallible. Although there are many factors that can improve memory performance, such as selecting appropriate memory strategies, people are poor at predicting what they will or will not remember and tend to choose strategies that are suboptimal or counterproductive. In our research, we are investigating patterns of brain activity associated with good and poor memory performance. We are examining methods for improving human performance by identifying cases where learners are using suboptimal memory strategies. Through this effort, we hope to lay the foundation for closing the loop between recording brain activity and using those recordings to augment performance.

As a part of this effort, one of our goals is to create a model of brain activity that can be used in a predictive fashion. Using brain activity recorded from participants who tried to memorize words under a variety of study and test conditions, we selected two conditions where the brain's response to stimuli should be similar across all

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 77–86, 2011.

participants. We used those two conditions to develop a computation model and then tested the model on a third condition in which participants' brain activity should depend on their choice of study strategies. The model was used to predict which participants were using a more effective memory strategy. We then tested the predictions of the model by assessing the participants' behavioral memory performance.

1.1 Metamemory and Memory Strategies

The term *metamemory* refers to a person's judgments about the state of his or her own memory. Successful encoding and retrieval of information requires a number of metamemory decisions, such as deciding what information is worth remembering, what strategies should be used to encode the information, and whether or not information retrieved from memory is accurate. People typically develop metamemory skills over time, through experience with different kinds of learning situations. For example, after practice with sequences of study and test questions, people tend to get better at predicting which items they will remember later and which items need additional study [1]. Through experience, people learn memory strategies such as spending more time studying items that seem difficult to remember or using different study strategies depending on when and how the information will need to be remembered. However, the strategies that learners develop are often affected by cognitive biases and may not be optimal. Numerous studies have shown that people often fail to use appropriate memory strategies [2,3,4,5,6].

One memory strategy that can improve performance is self-testing, or retrieval practice [7]. A common example of retrieval practice is studying with flashcards. If a language student studies new vocabulary by quizzing herself with flashcards, she will be more likely to remember the new words than if she skimmed over the words and their definitions in a textbook. Retrieval practice is beneficial because it gives learners experience with retrieving the needed information from memory. It also provides learners with a more accurate sense of what they do and do not remember. The effectiveness of retrieval practice increases as the practice becomes more difficult [3].

Although retrieval practice is a highly effective strategy, it is not a strategy that learners are likely to adopt on their own. Studying with this strategy can be frustrating because learners feel that they are performing poorly and progressing slowly. In reality, they are developing accurate assessments of how well they have learned the material. However, learners tend to prefer study strategies that make them feel successful at the time of study, even when those strategies are less effective in the long run [2,3].

1.2 Event-Related Potentials

When a person's brain activity is recorded using electroencephalography (EEG), different patterns of activity emerge for passive study and for active retrieval of information from memory. In EEG research, a participant's brain activity is recorded using sensors placed on his or her scalp. The EEG data provide an ongoing record of the brain's electrical activity with very high temporal resolution. To separate out the brain activity related to a particular type of processing, the EEG data are time-locked to the presentation of events of interest and events of the same type are averaged together. The averaging process should average out any ongoing processing that was not related to the experimental stimuli, leaving only the activity that was elicited by the events of interest. These averaged waveforms are called event-related potentials (ERPs).

Researchers have mapped the relationships between different ERP waveforms and different types of processing in the brain. The ERP component of interest in the present study is the late positive component (LPC). The LPC is thought to be related to explicit processing, such as the process of deliberately searching memory for a particular piece of information [8, 9, 10, 11, 12].

When a learner is presented with an item to study, the LPC elicited by that item will be small. However, when the learner is tested on an item and has to retrieve it from memory, the presentation of that item will elicit a large LPC. In the absence of an explicit memory test, a participant's self-induced retrieval practice should also produce a larger LPC. It is likely that use of retrieval practice as a memory strategy will be reflected in participants' brain activity during study.

1.3 Modeling Event-Related Potentials

The EEG data selected for modeling was taken from a study in which participants were presented with a list of words and asked to remember them for a later memory test. Some of the words were studied once, some words were studied twice, and some were studied once and then quizzed once during the study session. All of the words appeared again on a subsequent memory test, intermixed with an equal number of new words. We hypothesized that participants would have the worst memory for the words that were studied only once and the best memory for the words that were quizzed during the study sessions. The quizzes provide an opportunity for retrieval practice that should benefit subsequent memory performance.

For the words that were studied twice but not quizzed, we hypothesized that some participants would engage in retrieval practice on their own. Even though the words were not explicitly tested, participants might recognize them as previously studied words and retrieve the first presentation of the word from memory. This self-testing should benefit subsequent memory performance much like explicit testing. Since, as discussed above, most people are unlikely to adopt a strategy such as retrieval practice on their own, we expected that the average performance across all participants would be lower for the twice-studied items than for the quizzed items. However, we expected that a subset of the participants would use more effective memory strategies and would perform better on this condition than their peers.

The design of the EEG experiment allowed us to model each participant's brain activity in two "known" conditions: the first presentation of each studied word, which should not elicit an LPC, and the words that were quizzed during the study block, which should elicit a large LPC. We applied the model to ERPs from an unknown condition, the second presentation of repeated study words. The words in that condition should elicit an LPC only for the participants who engaged in retrieval practice. We used the model to classify the ERPs from the unknown condition as being more like passively studied words or more like explicitly tested words. We then tested the predictive power of the model by comparing the subsequent memory performance for participants in those two groups.

2 Experimental Methods

Participants. Twenty-four University of Illinois students participated in this study and were paid for their participation. Half of the participants were male and half were female. The average age of the participants was 21.

Materials. The materials used in the experiment consisted of 320 common nouns that served as study items, and 320 nouns that were matched in terms of length and frequency and served as new items at test. The average frequency was 57.6 for the study items and 50.9 for the new items; the average word length was 4.6 letters for both sets of words (frequency data was taken from the Kucera and Francis, 1967; norms included in Balota et al., 2002; a frequency value of zero was assumed for items not appearing in the database).

The study words were divided into eight counterbalanced lists. The experimental lists were subdivided into four study blocks and four test blocks. Each study block contained 80 of the experimental items. Of those items, 20 were studied once, 20 were studied twice, 20 were studied and then tested within the block, and 20 were paired with a synonym. For the items that were studied twice or studied and then tested, half of the items were repeated at a short lag, defined as one intervening item, and half were repeated at a long lag, defined as nine intervening items. For the items that were paired with synonyms, half of the synonyms were presented at a short lag and half were presented at a long lag. In addition, half of the synonym items were tested at each lag.

Each study block was followed by a test block in which all of the nouns from the block were re-tested, intermixed with an equal number of new, unstudied items.

Procedure. The participants were instructed that they would be tested on their memory for a list of study words. They were not given any information about different types of memory strategies and were not asked to use a particular memory strategy. As discussed above, the study list was broken into four parts in order to make the task easier for the participants. Each study block contained a total of 140 study words and each test block contained a total of 160 test words.

Throughout the experiment, there was a white fixation cross in the center of the computer screen. The participants were asked to keep their eyes on the fixation cross at all times during the experiment. All of the study words were presented immediately above the fixation cross in white 38-point Helvetica font on a black background. Within the study blocks, each word was preceded by a pound symbol (#) that was presented above the fixation cross for one second. Participants were instructed that they could blink or move their eyes while the pound symbol was on the screen, but that when it disappeared they should refrain from blinking and prepare to see the next study word. For the tested words, the pound symbol was red, indicating that the next word would be tested. For the words that were only studied, the pound symbol was white. The study word was presented 500 ms after the pound symbol disappeared and remained on the screen for one second. The tested words were followed by a red question mark that remained on the screen until the participants pressed a response button to indicate whether or not that word had appeared earlier in the study block. The same test procedure was used in the test blocks that followed each study block. In the test blocks, all of the words from the study block were tested or retested, intermixed with an equal number of new words. The participants took short breaks before starting each new study block in order to reduce interference from the preceding blocks.

The electroencephalogram (EEG) was recorded from 26 silver/silver-chloride electrodes embedded in a geodesic arrangement in an elastic cap (EASY-cap). Five additional free electrodes were placed on the left and right mastoids, on the outer canthus of each eye, and below the left eye. The three free electrodes near the eyes were used to record blinks and horizontal eye movements (vertical and horizontal EOG). The scalp electrodes were referenced on-line to the left mastoid. Following the experiment, the scalp electrodes were re-referenced off-line to an average of the left and right mastoids. All of the electrodes were tested before recording begins to ensure that their impedance was below 3 KOhms. During the experiment, the EEG from all electrodes was amplified through a bandpass filter of 0.02-100 Hz and recorded at a sampling rate of 250 Hz.

ERPs were computed at each electrode for each experimental condition by averaging the EEG data from 100 ms before the onset of a word until 920 ms after word onset. Trials containing blinks were corrected using the blink correction procedure described by Dale (1994) and trials containing artifacts such as excessive eye movement, signal drift or muscle activity were excluded from the averages. The mean amplitude of the ERPs within time windows of interest was calculated using data digitally filtered off-line using a bandpass filter of 0.2 to 20 Hz.

3 Experimental Results

Behavioral Results. Memory accuracy was assessed using the percentage of correct answers on the memory tests. Only the data relevant to the computational model will be discussed here. On average, participants were 39% correct for words that were studied only once, 50% correct for items that were studied twice with a long lag between the repetitions, and 66% correct for items that were studied once and quizzed at a long lag during the study block. These results were consistent with the prediction that retrieval practice during study would benefit subsequent memory performance. The difference in performance between the twice-studied words and the quizzed words also supports the hypothesis that most participants would not use retrieval practice when presented with repeated study words.

ERP Results. The LPC was measured by computing the mean amplitude of the ERPs in a time window from 500-900 ms post stimulus onset. Repeated measures ANOVAs were used to test the results, with degrees of freedom adjusted using the Greenhouse-Geisser correction. All effects are significant at or above the $p = 0.05$ level unless otherwise specified.

The LPC was significantly larger for the words that were quizzed during the study block than for those that were not, as shown in Figure 1. As predicted, this indicates that participants actively searched their memory for the words that were explicitly quizzed. However, for the words that were studied twice, most (if not all) of the participants studied the words passively. They did not search their memory to retrieve the previous presentation of the words, so the second presentation of the words did not elicit and LPC.

Although the majority of the participants did not engage in retrieval practice when they were not explicitly tested, we developed a model to identify whether or not there were any subgroups of participants who did employ that strategy.

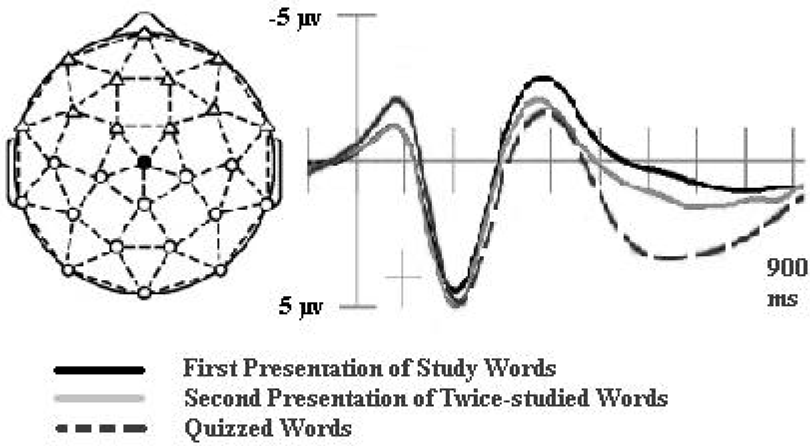

Fig. 1. Grand average ERPs to first presentation of studied words (black line), second presentation of twice-studied words (gray line), and quizzed words (dotted line). ERPs are shown at the midline central (MiCe) electrode.

4 Computational Modeling

Our goal was to construct a computational model that would classify ERPs elicited by the words in the twice-studied condition based on the brain activity associated with a particular study strategy (retrieval practice or passive study). This was achieved by constructing a naive Bayes classifier trained on the known study and test conditions and applying this classifier to the unknown ERPs. A significant challenge faced when constructing computational models from EEG signals is the low signal-to-noise ratio due to the presence of simultaneously recorded brain activity that is unrelated to the event of interest. This is often addressed by averaging all single-trial EEG recordings to form a grand average ERP. However this approach removes most of the trial-to-trial variability and can result in the formation of a classifier that is not robust to variances present in the ERPs from the "unknown" condition.

To overcome these obstacles, we developed an approach that better balances variability and signal averaging. Our approach combines ensembling classification results from multiple models and randomized signal averaging of individual trial ERPs. Randomized signal averaging was accomplished using an n-choose-k approach to create a new set of ERPs for use in the classifier training step. We examined maximized signal averaging by using k=39 to select and average single trial EEG recordings in a time window from 100 ms pre-stimulus to 900 ms post-stimulus to create 40 ERP samples for each of the two known conditions (study and test). For the study condition (the first presentation of all studied words), there were 278 single trial EEG recordings available from which to choose and for the test condition (the words that were quizzed during the study block) there were 40 single trial EEG recordings. The resulting ERP samples were then transformed via principal component analysis and the scores of the first five principal components were used as an uncorrelated feature set to train a naive Bayes classifier. The classifier was implemented by using MATLAB's [13] classify function provided in the Statistical Toolbox with the "diaglinear" discriminant function. This process was then repeated 50 times using a new random seed to randomize the single trial EEG recordings chosen for signal averaging from the n-choose-k trial selection process. In this way, each model was exposed to different signal averaging in the unknown condition ERP samples while maintaining a balanced number of training examples across the two known conditions.

5 Modeling Results

The performance of the classifier was estimated using sample-out cross validation. For this work, single trial data was available from twenty-three of the participating subjects.

The mean area under the receiver-operator curve (AUC) for sample-out cross validation over all models and all subjects was 0.99. The standard deviation of the mean sample-out cross validation AUC for each subject was 0.01. These cross validation results provide confidence that the feature extraction and classification methods are well suited to model the brain activity related to passive study or retrieval practice strategies.

For classification of the unknown ERPs, a full model was constructed with all samples from each of the 50 randomly constructed training sets described in section 4. This model was then used to classify the unknown ERPs as belonging to the study or retrieval groups. Examination of the number of models classifying the unknown ERPs as belonging to the study group identified eighteen subjects whose brain activity was consistent with their previously used study strategy. For this group of eighteen subjects, more than 97% of the models for each subject identified the unknown ERPs as belonging to the study class. Another group of five of the subjects exhibited brain activity that was consistent with the retrieval practice elicited by the quizzed words. The number of models identifying the unknown ERPs as belonging to the test class varied with subject and ranged from 22% to 80% of the 50 models constructed for each subject (Table 1). This variation is indicative of individual differences and may indicate that the retrieval practice strategy was employed with different frequency by each subject.

Table 1. Percentage of models indicating a retrieval practice strategy for each subject

Subject	Percentage of models indicating retrieval strategy	Percentage of twice-studied words remembered at test
27	80%	73%
15	78%	78%
2	54%	58%
12	52%	88%
3	22%	63%
22	2%	33%
5	0%	43%
7	0%	18%
8	0%	13%
9	0%	58%
10	0%	45%
11	0%	58%
13	0%	45%
14	0%	80%
16	0%	68%
18	0%	28%
19	0%	38%
20	0%	63%
21	0%	58%
26	0%	38%
28	0%	23%
29	0%	55%
30	0%	58%

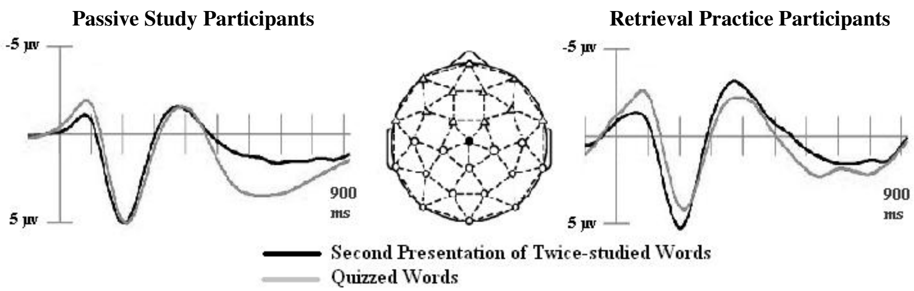

Fig. 2. Grand average ERPs to the unknown condition, the *second presentation of twice-studied words (black line)*, and the test condition, the *quizzed words (gray line)*. ERPs are shown at the midline central (MiCe) electrode. The participants whose brain activity was consistent with passive study in the unknown condition are shown on the left and the participants whose brain activity was consistent with a retrieval practice strategy are shown on the right.

To test the model's classification performance, we compared the behavioral memory performance across the two groups of participants. As predicted, the participants whose brain activity in the unknown condition was consistent with their brain activity in the retrieval practice condition had better memory for the twice-studied items than participants whose brain activity was consistent with passive study. On average, the participants in the former group correctly recognized 28.6 out of 40 words (71.5%) from the twice-studied condition, while the participants in the latter group correctly recognized 18.1 out of 40 words (45.2%). Welch's t-test showed that the performance of the two groups was significantly different [$t(9.4) = 3.82$, $p < 0.01$]. Figure 2 shows the grand average ERPs for the unknown condition and the test condition for the two groups.

6 Discussion

The results of this experiment indicate that ERPs elicited under known conditions can be modeled and used to classify ERPs from an unknown condition. In this experiment, the known conditions included a passive study condition and a condition in which participants were quizzed on previously studied words, leading the participants to engage in retrieval practice. The unknown condition was the second presentation of repeated study items. For those items, participants might retrieve the first presentation of the word from memory, adopting a retrieval practice strategy on their own. Previous research on study strategies and cognitive biases led us to predict that few participants would spontaneously engage in retrieval practice, but those that did would outperform the other participants for the words in that condition.

As we predicted, the average memory performance across all participants was lower for the words that were studied twice than for the words that were studied and then quizzed. Using the model, we identified a group of five participants whose brain activity was consistent with use of a retrieval practice strategy. That small subset of participants performed significantly better than the other participants on the subsequent memory test.

The experiment and model described in this paper represent the first steps toward using recorded brain activity to improve human memory performance. We have identified patterns of brain activity that are associated with the use of an effective memory strategy and developed a model that can predict which participants are using that strategy and which are not. In future research, we hope to expand on these results and investigate ways to coach people on the effectiveness of their study strategies as they attempt to learn new information.

Acknowledgments. The authors would like to thank Dr. Kara Federmeier for her invaluable advice and assistance with this project. We also thank Sean Cusick, Bridget Milligan, and Charlotte Laguna for their assistance with data collection. This work was supported by the Laboratory Directed Research and Development (LDRD) Program at Sandia National Laboratories. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

References

1. Benjamin, A.S.: Memory is more than just remembering: Strategic control of encoding, accessing memory, and making decisions. In: Ross, B.H., Benjamin, A.S. (eds.) The Psychology of Learning and Motivation: Skill and Strategy in Memory Use, vol. 48, pp. 175–223. Academic Press, London (2008)
2. Benjamin, A.S., Bjork, R.A., Schwartz, B.L.: The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. Journal of Experimental Psychology: General 127, 55–68 (1998)
3. Bjork, R.A.: Assessing our own competence: Heuristics and Illusions. In: Gopher, D., Koriat, A. (eds.) Attention and Performance XVII: Cognitive Regulation of Performance: Interaction of Theory and Application, pp. 435–459. MIT Press, Cambridge (1999)
4. Koriat, A., Bjork, R.A.: Illusions of competence in monitoring one's knowledge during study. Journal of Experimental Psychology: Learning, Memory, and Cognition 31, 187–194 (2005)
5. Koriat, A., Bjork, R.A., Sheffer, L., Bar, S.K.: Predicting one's own forgetting: The role of experience-based and theory-based processes. Journal of Experimental Psychology: General 133, 643–656 (2004)
6. Schwartz, B.L., Benjamin, A.S., Bjork, R.A.: The inferential and experiential bases of metamemory. Current Directions in Psychological Science 6, 132–137 (1997)
7. Landauer, T.K., Bjork, R.A.: Optimum rehearsal patterns and name learning. In: Gruneberg, M.M., Morris, P.E., Sykes, R.N. (eds.) Practical Aspects of Memory, pp. 625–632. Academic Press, London (1978)
8. Allan, K., Rugg, M.D.: An event-related potential study of explicit memory on tests of cued recall and recognition. Neuropsychologia 35, 387–397 (1997)
9. Neville, H.J., Kutas, M., Chesney, G., Schmidt, A.: Event-related brain potentials during the initial encoding and subsequent recognition memory of congruous and incongruous words. Journal of Memory and Language 25, 75–92 (1986)
10. Paller, K.A., Kutas, M.: Brain potentials during memory retrieval provide neuropsychological support for the distinction between conscious recollection and priming. Journal of Cognitive Neuroscience 4, 375–391 (1992)
11. Rugg, M.D., Doyle, M.C.: Event-related potentials and stimulus repetition in direct and indirect tests of memory. In: Heinze, H.J., Munte, T.F., Mangun, G.R. (eds.) Cognitive Electrophysiology, pp. 124–138. Birkhauser, Boston (1994)
12. Van Petten, C., Senkfor, A.J.: Memory for words and novel visual patterns: Repetition, recognition, and encoding effects in the event-related brain potential. Psychophysiology 33, 491–508 (1996)
13. MATLAB version 7.11.0.584 (R2010b). The MathWorks Inc., Natick (2010)