Late Positive Component Event-related Potential Amplitude Predicts Long-term Classroom-based Learning

Katherine W. Turk\textsuperscript{1,2}, Ala’a A. Elshaar\textsuperscript{1,2}, Rebecca G. Deason\textsuperscript{3}, Nadine C. Heyworth\textsuperscript{4}, Corrine Nagle\textsuperscript{1,2}, Bruno Frustace\textsuperscript{1,2}, Sean Flannery\textsuperscript{1,2}, Ann Zumwalt\textsuperscript{2}, and Andrew E. Budson\textsuperscript{1,2}

Abstract

It is difficult to predict whether newly learned information will be retrievable in the future. A biomarker of long-lasting learning, capable of predicting an individual's future ability to retrieve a particular memory, could positively influence teaching and educational methods. ERPs were investigated as a potential biomarker of long-lasting learning. Prior ERP studies have supported a dual-process model of recognition memory that categorizes recollection and familiarity as distinct memorial processes with distinct ERP correlates. The late positive component is thought to underlie conscious recollection and the frontal N400 signal is thought to reflect familiarity [Yonelinas, A. P. Components of episodic memory: The contribution of recollection and familiarity. Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 356, 1363–1374, 2001]. Here we show that the magnitude of the late positive component, soon after initial learning, is predictive of subsequent recollection of anatomical terms among medical students 6 months later.

INTRODUCTION

The commitment of newly learned information to long-term memory storage is important for all organisms, allowing past experiences and knowledge to influence future decisions. In the medical professions, long-term learning is vital, as daily decisions are often based on knowledge acquired years earlier, during medical school training. It is currently difficult to determine at the time of learning whether particular teaching methods are efficacious in helping students encode, consolidate, and later retrieve knowledge from long-term memory storage. A biomarker of long-lasting learning that could predict students' ability to consolidate and maintain novel information in long-term memory storage would therefore be highly valuable to educators and society. These issues motivated the search for such a marker using ERPs, a time-locked measure of electrophysiological voltage changes in the brain, given their millisecond time resolution, are an ideal method for measuring subsequent memory, reliably linking neural activity elicited by a particular stimulus to later memory for the event [Rugg & Curran, 2007]. Others have previously shown that remembered items elicit more positive ERP waveforms during recognition than either new or forgotten items (Woodruff, Hayama, & Rugg, 2006). The dual-process model of recognition memory states that recollection, the ability to recognize a stimulus by defining item-specific characteristics, and familiarity, more general awareness of a prior exposure to a stimulus, are distinct memorial processes with separate neural substrates (Yonelinas, 2001). Two separate ERP correlates of recognition have been identified with different latencies and scalp distributions, the late positive component (LPC) and the early frontal N400 signal (FN400). In standard recognition memory experiments, the LPC is observed to occur initially around 600 msec poststimulus, peaking by 1000 msec over the left parietal lobe, and has been theorized to represent the ERP correlate of conscious recollection (Yu & Rugg, 2010; Rugg & Curran, 2007; Woodruff et al., 2006; Curran & Cleary, 2003; Rugg & Yonelinas, 2003; Duzel, Yonelinas, Mangun, Heinze, & Tulving, 1997). The FN400 is maximal frontally, occurs around 300–500 msec poststimulus, and may reflect familiarity (Yu & Rugg, 2010; Rugg & Curran, 2007; Woodruff et al., 2006; Curran & Cleary, 2003; Rugg & Yonelinas, 2003). We therefore hypothesized that the amplitude of the LPC, following learning of novel words, may predict the learner’s ability to accurately recollect and define words up to 6 months later, indicating effective long-term consolidation, storage, and retrieval of information.

\textsuperscript{1}VA Boston Healthcare System, \textsuperscript{2}Boston University School of Medicine, \textsuperscript{3}Texas State University, \textsuperscript{4}University of California, San Diego

© 2018 Massachusetts Institute of Technology

Journal of Cognitive Neuroscience 30:9, pp. 1323–1329
doi:10.1162/jocn_a_01285

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/jocn_a_01285 by guest on 28 August 2021
ERPs have been utilized in studies investigating subsequent memory performance (Otten, Quayle, Akram, Ditewig, & Rugg, 2006; Sanquist, Rohrbaugh, Syndulko, & Lindsley, 1980). Our study, however, uses a distinct paradigm to extend prior findings to real-life, course-based learning of novel words. Medical students enrolled in an introductory anatomy course were presented with current as well as antiquated anatomical terms during ERP recording at three time points: before the course, immediately after the course end date (16 weeks later), and finally 6 months after the course end date. Participants were asked to rate their knowledge of the presented terms using the responses Can Define (CD), Familiar (Fam), or Don’t Know (DK). Others have shown that recollection tends to degrade more quickly than familiarity (Gardiner & Java, 1991). It was hypothesized that the Fam responses would increase over time across the sessions showing sustained levels of familiarity and the proportion of CD terms would be greatest at Session 2 and then decrease at Session 3, indicating decreases in recollection over time. To look for a potential biomarker of long-lasting memory, we examined ERP responses at Session 2 separated according to subsequent memory performance 6 months later at Session 3.

METHODS

Thirty-four right-handed adults (12 men, 22 women), aged 20–30 years (mean age = 23.97 years, SD = 2.53 years), were recruited for this ERP study. Participants were first-year medical students enrolled at Boston University School of Medicine who were proficient in the English language and who had not previously taken a formal anatomy course. All participants were healthy and had normal or corrected-to-normal vision. Participants completed three sessions and were compensated $10 per hour for their participation.

A total of 264 anatomical terms were presented over the course of the three experimental sessions. A total of 132 relevant terms were selected from the learning objectives of the anatomy course, 44 each from the three sections: Back and Limbs, Thorax, Abdomen and Pelvis, and Head and Neck. There were also 132 obscure terms that were antiquated anatomical terms of Latin and Arabic roots, which should not have been familiar to the students (e.g., alagmur, adnexus; taken from Fonahn, 1922). Each of the three experimental sessions consisted of the same 132 relevant terms and 44 different obscure terms for a total of 176 words. The 44 obscure terms were new in each of the three experimental sessions so that there was no overlap between sessions. Session 1 occurred at baseline, before the start of the anatomy course; Session 2 occurred immediately after the course end (16 weeks after Session 1); and Session 3 occurred 6 months following Session 2.

During each session, participants were presented with 176 terms randomly presented on a 15-in. Dell computer screen using E-Prime software, Version 2.0 (Schneider, Eschman, & Zuccolotto, 2002). Participants were told that they would see a series of anatomical terms and were instructed to make one of three different decisions by keystroke for each term: “Can Define” if they are able to define the structure, function, or location of the term if asked; “Familiar” if they recognized the term but could not define the structure, function, or location of the term; or “Don’t Know” if they are not familiar with the term. All decisions were self-paced. Participants were informed that they would complete a posttest at the conclusion of the experiment to ensure that “Can Define” responses reflected the participants’ true ability to recollect the anatomical terms. Each trial started with a 1500-msec ISI followed by the presentation of the anatomical term in the center of screen. After the participant responded, the next anatomical term would appear after a 1500-msec ISI.

A posttest was created based on recorded behavioral responses. Participants were presented with the words they marked as “Can Define” during the experimental procedure and were asked to generate the appropriate definition on an Excel spreadsheet. Posttests were examined for accuracy, and participants that did not complete or earned less than 50% on the posttest were excluded from analyses. Of the 60 enrolled participants, 49 completed the ERP testing, 15 of which had unusable data. In 11 of these cases, the number of responses in at least one of the categories (CD, Fam, or DK) was too small to allow creation of meaningful ERP waveforms due to signal-to-noise issues. Two of the participants, although they had adequate responses, had unusable ERP data due to corrupt EEG files that were unretrievable. Lastly, there were two participants who had unusable data due to high levels of artifact. In total, 34 participants had usable ERP data included in the analysis.

Participants were seated in a hardback chair and fitted with an Active Two electrode cap (Behavioral Brain Sciences Center). A fill array of 128 Ag-AgCl BioSemi Active electrodes was connected to the cap in a preconfigured montage, which places each electrode in equivalent concentric circles from 10–20 position, Cz. In addition to the 128 scalp electrodes, mini-biopotential electrodes were places on each mastoid process. Finally, vertical and horizontal EOG activity was recorded from bipolar electrodes placed below the left eye and on the outer canthus of the left and right eyes. EEG and EOG activities were amplified with a bandwidth of 0.03–35 Hz (3 dB points) and digitized at a sampling rate of 256 Hz. Recordings were referenced to a vertex point but were later re-referenced to a common average reference to minimize the effects of reference site activity and to accurately estimate the scalp topography of the measured electrical fields (Ally & Budson, 2007). The sampling epoch for each test trial lasted for a total of 2200 msec, which included a 200-msec prestimulus baseline period. This prestimulus period was used to

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/jocn_a_01285 by guest on 28 August 2021
baseline-correct averaged ERP epochs lasting 2000 msec. ERPs were averaged and corrected using the EMSE Software Suite (Source Signal Imaging). Trials were corrected for excessive EOG activity using the EMSE Ocular Artifact Correction Tool. The tool first allows the investigator to manually distinguish artifact data from artifact-free data. Then, using a covariance technique that simultaneously models artifact and artifact-free EEG, a logarithmic ratio of artifact data to clean data is produced by EMSE. Finally, ocular artifact is subtracted from the recording where it is detected by the correction tool. Trials were discarded from the analyses if they contained baseline drift or movement greater than 90 V. Individual bad channels were corrected with the EMSE spatial interpolation filter.

Nine ROIs were determined in advance (Figure 1), and the left parietal region, ROI 6, was used for all ERP analyses here based on a priori assumptions. Proportions of each behavioral response type per participant were derived by dividing the total number of a single response type (CD, Fam, DK) by the total number of responses to current anatomical terms (132).

Mean amplitudes were calculated from time intervals of 700–1000 msec (after stimulus onset), which were then averaged across a group of relevant electrodes in the left parietal region. A univariate repeated-measures ANOVA was performed comparing the effect of response type at Session 3 (CD, Fam, and DK), at ROI 6 (based on a priori expected outcomes), for grand-averaged ERP data from Session 2, 700–1000 msec poststimulus. Statistical analyses were performed using statistical software (SPSS Version 16.0, SPSS Inc.). Two-tailed $p$ values of less than or equal to .05 were considered significant. The Greenhouse–Geisser correction was applied where appropriate to correct for violations of sphericity. The waveforms and scalp topographies were formed by averaging a series of trials for each participant.

**RESULTS**

Proportion of CD, Fam, and DK responses among relevant terms only was measured across the three sessions (Figure 2). Using a Session × Response type repeated-measures ANOVA, a main effect of Response type was observed, $F(2, 66) = 11.08, p < .001, \eta_p^2 = .25$. An effect of session was also observed, $F(4, 132) = 283.07, p < .001, \eta_p^2 = .90$, indicating that the proportion of responses was different between sessions. A repeated-measures ANOVA of CD responses revealed a change in responses across sessions, $F(2, 66) = 236.91, p < .001, \eta_p^2 = .88$. Post hoc pairwise comparisons revealed an increase in CD responses from Session 1 to Sessions 2 and 3 (0.05 ± 0.05 vs. 0.64 ± 0.15 vs. 0.50 ± 0.17, both at $p < .001$). CD responses peaked at Session 2 and showed a decline in Session 3 ($p < .001$). An ANOVA of Fam responses across sessions revealed differences, $F(2, 66) = 35.79, p < .001,$
Post hoc tests showed an increase in Fam responses between Sessions 1, 2, and 3 (0.17 ± 0.11 vs. 0.27 ± 0.11 vs. 0.40 ± 0.14, respectively, ps < .001). An ANOVA of DK responses also differed across sessions, $F(2, 66) = 939.94, p < .001$, $\eta^2_p = .97$. Post hoc tests showed a decrease in DK responses from Session 1 to Session 2 (0.78 ± 0.14 vs. 0.09 ± 0.06, $p < .001$) and Session 1 to Session 3 (0.78 ± 0.14 vs. 0.10 ± 0.07, $p < .001$).

When individual CD terms from Session 2 were evaluated as to whether they became CD, Fam, or DK responses in Session 3, results revealed that 60.5% of CD responses remained so, 34.4% of responses became Fam, and 5.1% of responses became DK. Similar proportions were found when confirmed CD terms were analyzed from Session 2: 65.0% of confirmed CD responses remained so at Session 3, 31.3% of responses became Fam, and 3.7% of responses became DK.

ERP waveforms generated during Session 2 following exposure to current anatomical terms only were grouped by participants’ CD, Fam, and DK behavioral responses from Session 3. Similar to prior recognition ERP studies, a left parietal LPC was observed that differed between response types between 700 and 1000 msec after stimulus presentation (Figures 3 and 4). Based on these a priori

---

Figure 2. Proportion of Can Define and Familiar behavioral responses to relevant terms categorized by session. In each case, the remaining proportion of responses are composed of Don’t Know responses. The proportion of both Can Define and Familiar responses increased from Session 1 to Session 2, and the proportion of terms which the participant was familiar with but unable to define increased in Session 3 from Session 2. Error bars show SEM.

Figure 3. Session 2, left posterior central region ERP grand-averaged waveforms across all 34 participants separated by behavioral response at Session 3. Period of interest, 700–1000 msec poststimulus, used in statistical analysis.
expected outcomes, a repeated-measures ANOVA for the left parietal region between Session 3 Response type and average Session 2 ERP amplitude from 700 to 1000 msec revealed an effect of Response type, $F(2, 33) = 20.07, p < .001, \eta^2_p = .38$. Post hoc tests revealed increased amplitude of the LPC for CD versus DK ($p < .001$) and for Fam versus DK ($p < .001$) and CD versus Fam ($p < .05$). We also performed a follow-up post hoc analysis with a subset of the CD words that were confirmed as correctly defined by the Session 3 posttest using Session 2 ERP waveforms and found the persistence of a strong LPC effect at the same time interval (700–1000 msec).

The FN400 amplitude from the right frontal ROI using a 300–500 msec time interval paired with Session 3 CD, Fam, and DK responses was evaluated using a repeated-measures ANOVA that revealed an effect of Response type of $F(2, 33) = 11.22, p = .0003, \eta^2_p = .25$. Post hoc tests revealed no difference between the FN400 amplitude for CD versus Fam, $p = .246$, whereas there was a significant difference between CD versus DK, $p = .003$, and Fam versus DK, $p = .0002$. This analysis revealed that the FN400 amplitude at Session 2 was able to distinguish between both CD and DK, and Familiar and DK responses, but not between CD and Fam responses.

**DISCUSSION**

Our behavioral results indicate that the majority of information learned and recollected as definable terms during Session 2 were preserved 6 months later. As predicted, some recollected information degraded over time and became terms that were merely familiar, with the defining details no longer retrievable. Our ERP results revealed a neural correlate associated with the long-term recollection of terms learned during an anatomy course. This ERP effect presented as an increase in LPC amplitude occurring most robustly 700–1000 msec poststimulus over the left parietal scalp. Increased positivity of this late-positive ERP signal was able to distinguish between stimuli that would be recollected up to 6 months later versus those stimuli recognized as familiar, but lacking specific defining details, as well as those stimuli that were completely forgotten.

Whereas the LPC amplitude at Session 2 was found to be predictive of later CD responses, the FN400 at Session 2, a frontal ERP signal thought to be associated with familiarity (Curran & Cleary, 2003), was able to distinguish between later Fam versus DK responses and CD versus DK response pairs at Session 3. The FN400 amplitude at Session 2 was not associated with later CD versus Fam responses during Session 3. These findings fit with our hypothesis that LPC alone would be predictive of later CD versus Fam responses as a potential marker of high-quality, long-lasting learning. FN400 may be acting as a marker of more general familiarity, the ability to distinguish a term as either unknown (DK) or known (Fam or CD).

Interestingly, we also found that individual terms in the CD category from Session 2 remain as CD terms in nearly two thirds of cases at Session 3, become merely familiar in a little less than one-third of cases, and only become unknown terms with DK responses about 3% of the time. This result suggests that the behavioral response at Session 2 may hold some predictive power about the likelihood of long-term memory formation as well.

Others have used a subsequent memory paradigm examining verbal encoding of distinctive words and found an LPC subsequent memory effect, occurring maximally centroparietally 800 msec after stimulus onset and lasting 600–1300 msec (Fernandez et al., 1998). The LPC occurred somewhat later for low-frequency words that participants were less likely to encounter previously, similar to our findings here of an LPC effect occurring 700–1000 msec for newly learned anatomical terms.

It has been suggested that the neural bases of components of recognition memory differ with the hippocampus primarily involved with recollection and the adjacent perirhinal cortex involved with familiarity (Eichenbaum, Yonelinas, & Ranganath, 2007; Brown & Aggleton, 2001). Imaging studies have demonstrated that the hippocampus is involved in encoding and retrieval of recognition memories when they are strong (Rugg et al., 2012; Smith, Wixted, & Squire, 2011). Evidence for the hippocampus as a neural generator of the LPC subsequent memory effect arises from lesion studies showing that LPC positivity is decreased or absent in patients with temporal lobe lesions involving the hippocampus (Rugg, Roberts, Potter, Pickles, & Nagy, 1991; Smith & Halgren, 1989). Hippocampal–parietal networks may be an additional neural generator of the LPC. Neuroimaging studies have reported medial and lateral parietal activation with memory retrieval tasks (Wagner, Shannon, Kahn, & Buckner, 2005; Rugg, Otten, & Henson, 2002). Other fMRI studies have shown greater parietal activation for items participants reported recollecting
strongly versus being familiar with (Wagner et al., 2005; Shannon & Buckner, 2004; Eldridge, Knowlton, Furmanski, Bookheimer, & Engel, 2000), similar to our CD/Fam paradigm. The LPC may represent conscious awareness of a retrieved memory as processed by the parietal cortex and its hippocampal connection (Ally, Simons, McKeever, Peers, & Budson, 2008).

Our study provides new knowledge about the LPC as a result of the distinct, course-based learning paradigm used. First, participants in the study had little prior exposure to the anatomical words being learned, perhaps most similar to a second-language vocabulary acquisition paradigm, in which participants are asked to learn entirely novel words (Palmer, Havelka, & van Hooft, 2013). In addition, most prior studies of the ERP subsequent memory effect have involved experimental designs that measure ERPs at time of encoding with retrieval occurring several minutes to several weeks later (Tsivilis et al., 2015; Wolk et al., 2006). Rather than examining encoding and subsequent retrieval as in a traditional ERP subsequent memory design, our study used a classroom setting for a more naturalistic encoding environment in which participants were presented with terms in multiple settings over the course of months. We then examined changes in recognition memory across recognition tests with a 6-month retention interval to measure subsequent memory, allowing additional understanding of the ERP signals associated with long-term recollection.

There is no universal agreement as to the best measure of educational knowledge retention. The most commonly used measures in educational research are cue recall and recognition memory (Custers, 2010). Prior behavioral studies of the rates of retention of medical student knowledge and classroom-based learning more generally (D’Eon, 2006; Blunt & Blizard, 1975) have not examined the depth of prior knowledge stores and whether knowledge learned in courses is merely familiar or whether it evokes a deeper sense of recollection. Here we present a new finding that a more positive LPC ERP component measured at the end of a course may reflect future learning success and memorial strength several months later. This result has the potential to advance the examination of educational knowledge retention, moving from traditional behavioral studies of test results to physiological biomarkers of long-lasting learning.

These findings also have relevance for educational curriculum development. Our results allow for various teaching methods to be trialed in a classroom setting and the positivity of the LPC measured immediately at the end of the course—possibly even at the end of a particular lesson. Teaching methods that produce the greatest LPC positivity could then be further refined and tested, facilitating cycles of rapid educational improvement, enabling the development of novel teaching techniques that may engender the most robust and long-lasting learning.

Acknowledgments

This research was supported by Veterans Administration Merit Review award 5I01CX000736-03 and National Institute on Aging grant P30 AG013846.

Reprint requests should be sent to Katherine Turk, Center for Translational Cognitive Neuroscience, VA Boston Healthcare System, Boston, MA 02130, or via e-mail: kturk@bu.edu.

REFERENCES

D’Eon, M. F. (2006). Knowledge loss of medical students on first year basic science courses at the University of Saskatchewan. BMC Medical Education, 6, 5.
neuroimaging, Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 357, 1097–1110.


