

The Effects of Probe Similarity on Retrieval and Comparison Processes in Associative Recognition

Qiong Zhang, Matthew M. Walsh, and John R. Anderson

Abstract

■ In this study, we investigated the information processing stages underlying associative recognition. We recorded EEG data while participants performed a task that involved deciding whether a probe word triple-matched any previously studied triple. We varied the similarity between probes and studied triples. According to a model of associative recognition developed in the Adaptive Control of Thought-Rational cognitive architecture, probe similarity affects the duration of the retrieval stage: Retrieval is fastest when the probe is similar to a studied triple. This effect may be obscured, however, by the duration of the comparison stage, which is fastest when the

probe is not similar to the retrieved triple. Owing to the opposing effects of probe similarity on retrieval and comparison, overall RTs provide little information about each stage's duration. As such, we evaluated the model using a novel approach that decomposes the EEG signal into a sequence of latent states and provides information about the durations of the underlying information processing stages. The approach uses a hidden semi-Markov model (HSMM) to identify brief sinusoidal peaks (called bumps) that mark the onsets of distinct cognitive stages. The analysis confirmed that probe type has opposite effects on retrieval and comparison stages. ■

INTRODUCTION

A longstanding interest in cognitive science is to identify the number and durations of different information processing stages involved in task performance (Donders, 1969, translation). The challenge is to identify the number of stages, measure their durations, and understand how different experimental factors affect those durations. Sternberg (1969) proposed the additive factor method to deal with these challenges. The method entails the following assumptions: (1) time between a stimulus and response is occupied by a stream of successive processing stages, (2) each stage begins after the preceding stage ends, (3) different experimental factors affect the durations of different stages, and (4) the effects of different values of an experimental factor can be seen in overall RTs. A limitation of the additive factor method is that if one experimental factor affects the durations of multiple stages, its impact on the duration of each stage cannot be easily identified. Worse yet, if a factor increases the duration of one stage while decreasing the duration of another, different levels of the factor may produce near-equivalent overall RTs. The cumulative nature of RTs limits conclusions that can be made about the effects of an experimental factor on the durations of individual processing stages.

In this article, we examine a case where an experimental factor is hypothesized to impact two processing stages in an associative recognition task in opposing ways. The fac-

tor (i.e., probe similarity) is the degree of match between the three words of a triple presented to participants (i.e., probe) and the words in a previously studied triple. Participants were asked to decide if they had studied the triple before. As described below, our theory of associative recognition predicts that similarity of the probe to a studied triple will decrease the duration of a retrieval stage while increasing the duration of a comparison stage. Consequently, the theory predicts small effects of similarity on overall RTs despite its larger effects on the durations of individual processing stages.

To overcome limitations of overall RT, we apply a non RT-based method that uses neuroimaging data gathered using EEG. The method involves applying hidden semi-Markov models and multivariate pattern analysis (HSMM-MVPA) to the EEG data to identify latent processing stages (Anderson, Zhang, Borst, & Walsh, 2016; Anderson & Fincham, 2014a, 2014b). Information about the number and durations of processing stages based on the EEG can be used to evaluate the predictions of an existing theory or to guide the development of a new theory. We use the HSMM-MVPA method to test our theory of how probe similarity affects associative recognition. In doing so, we advance understanding of associative recognition and demonstrate the utility of using the HSMM-MVPA method.

Associative Recognition

Associative recognition involves judging whether two or more items were previously encountered together. For

example, participants in our experiment decided whether three words in a triple had been studied together. According to one class of recall-to-reject models, such judgments would be made by retrieving a studied item from memory and comparing it to the probe to determine whether they match (Malmberg, 2008; Rotello & Heit, 2000; Rotello, Macmillan, & Van Tassel, 2000; Anderson & Reder, 1999). One example of a recall-to-reject model is the process level account based on Adaptive Control of Thought-Rational (ACT-R; Anderson, 2007). The ACT-R model consists of four general processing stages (Anderson et al., 2016; Borst, Schneider, Walsh, & Anderson; 2013; Schneider & Anderson, 2012): encoding the word in the probe, retrieving a related memory, comparing the retrieved item to the probe, and responding.

According to the ACT-R's theory of declarative memory (Anderson, 2007), the time to retrieve a memory is an inverse function of its activation, A_i

$$T_i = Fe^{-A_i}. \quad (1)$$

where F is a latency scaling parameter. The activation of an item is a sum of the item's inherent strength and its strength of associations with items present in the current context:

$$A_i = B_i + \sum_{j \in C} W_j S_{j,i}. \quad (2)$$

where B_i is its base level activation, C is the context defined as the set of retrieval cues, W_j is the attentional weight assigned to each cue j , and $S_{j,i}$ is the strength of the association between each cue j and item i . During associative recognition, words in the probe act as cues for retrieving a studied associate. The strength of association $S_{i,j}$ can be understood as the probability that cue j predicts item i . $S_{i,j}$ is expressed in ACT-R's associative strength equation:

$$S_{j,i} = S - \ln(\text{fan}_j). \quad (3)$$

S is a cue's maximum associative strength, and fan_j is the number of associates of cue j . In our experiment, we manipulated fan by varying the number of triples in which certain words appeared; the more triples a word appeared in, the less effective the word was as a retrieval cue. Many behavioral experiments have confirmed that RTs become longer as the number of associates or fan of a probe increases (e.g., Pirolli & Anderson, 1985; Anderson, 1974; for reviews, see Anderson, 2007; Anderson & Reder, 1999).

During associative recognition, ACT-R rejects foils (i.e., nonstudied associates) by retrieving the closest matching studied associate and determining that it does not perfectly match the probe. The retrieved associates in our experiment share one or more words with the probe. The time to retrieve the closest nonmatching associate to a foil will be longer than the time to retrieve the matching associate to a target (i.e., a studied associate). This is

because activation from the words in the foil spreads to different memories, whereas activation from the words in a target converges on a single memory. In other words, the number of sources spreading activation to the retrieved associate is greater for targets than for foils. This prediction, though straightforward, is difficult to test. The durations of other, nonretrieval-related processes may also vary, obscuring the effects of number of sources of spreading activation. In particular, the duration of the subsequent comparison stage in the retrieve-to-reject model may be longer when the retrieved item is similar to the probe.

Neuroimaging studies provide additional evidence for ACT-R's activation and associative strength equations. In two such studies (Danker, Gunn, & Anderson, 2008; Sohn et al., 2005), participants memorized word triples. They were then shown a set of probes and asked to decide whether they had studied the probes. The associative fans of words that made up the triples varied (Fans 1, 2, and 3). Consistent with behavioral studies, RTs increased with associative fan. Additionally, the fMRI BOLD response in the left pFC, which is postulate to reflect ACT-R's retrieval module, increased with associative fan. This supports the idea that the most demanding retrievals engaged the left pFC for the longest, producing the greatest BOLD response. Danker (2010) further found that foils produced greater LIPFC BOLD response, consistent with the proposal that they have the longest retrievals. Although these results are consistent with the ACT-R theory of associative recognition, they cannot be strongly tied to a retrieval stage because of fMRI's low temporal resolution.

EEG, unlike fMRI, has millisecond resolution, making it an attractive alternative for studying brief cognitive processes like those involved in associate recognition. EEG experiments of recognition memory have mainly focused on two ERP components: an early frontocentral negativity called the FN400 (Walsh, Paynter, Zhang, & Reder, 2016; Curran, 2000) and a later posterior positivity called the parietal old/new effect (Curran, 2000; Düzel, Yonelinas, Mangun, Heinze, & Tulving, 1997). These components have typically been interpreted in the context of dual-process theories of recollection memory (Rugg & Curran, 2007; Diana, Reder, Arndt, & Park, 2006; Yonelinas, 2002). The FN400 is thought to reflect a familiarity process that provides information about whether an item has been seen before but does not involve retrieval of contextual information (Curran, 2000). The parietal old/new effect corresponds to a recollection process and does involve retrieval of contextual information.

Given their role in recognition memory, one would think that these components would be informative with respect to models of associative recognition. This is especially true of the parietal old/new effect, which involves the retrieval of associative information. However, the conventional approach to isolating ERP components requires aligning the EEG signal to experiment events like stimulus onset or response commission and averaging

the signal across trials to create ERPs. However, differences in the onset latencies of components between trials and conditions can obscure or create the illusion of differences in ERP component amplitudes (Luck, 2005). This is especially problematic of components that are only weakly locked to overt experiment events.

HSMM-MVPA Applied to an EEG Data of Associative Recognition Task

To cope with the trial-by-trial variability in the onset latencies of ERP components, we developed a novel method that involves applying HSMM-MVPA to EEG data (Anderson et al., 2016). Transitions from one cognitive stage to the next are signified by the onsets of bumps that summate with ongoing sinusoidal noise in the EEG signal. The bumps have finite durations, amplitudes, and topographical distributions. The postulation that processing stages are signaled by such bumps is consistent with theories of ERP generation in EEG data (Yeung, Bogacz, Holroyd, & Cohen, 2004; Makeig et al., 2002). Using HSMM-MVPA, it is possible to recover the number, timing, and topographical distributions of bumps that maximize the likelihood of the EEG data.

We previously applied HSMM-MVPA to an EEG study of associative recognition that manipulated fan and probe type (Anderson et al., 2016). Participants saw targets made up of two words previously studied together and foils made up of two words previously studied in separate pairs. Figure 1 shows the swimlane representation of the ACT-R retrieve-to-reject model for targets and foils in that experiment, with module activities initiated by production rules. To elaborate how the EEG signal is mapped to different processing stages within the framework of the ACT-R theory, a production evokes a change in neural processing, which produces a phasic response characterized by a bump in the EEG signal. The HSMM-MVPA method identifies both the distinctive scalp profiles and the variable latencies of such bumps in the single-trial EEG. Two early bumps marked the encoding of the word pair, and a third bump marked the retrieval onset. Following a variable period, a fourth bump marked the completion of retrieval and the start of comparison. A final, fifth bump marked completion of comparing the retrieved and target word pairs with a behavioral response. Several of these bumps related to ERP components evoked in standard recognition memory paradigms. In

particular, during the time surrounding the fifth bump, voltages were more positive over the posterior scalp for targets than foils, corresponding to the parietal old/new effect.

Motivation and Overview of Current Experiment

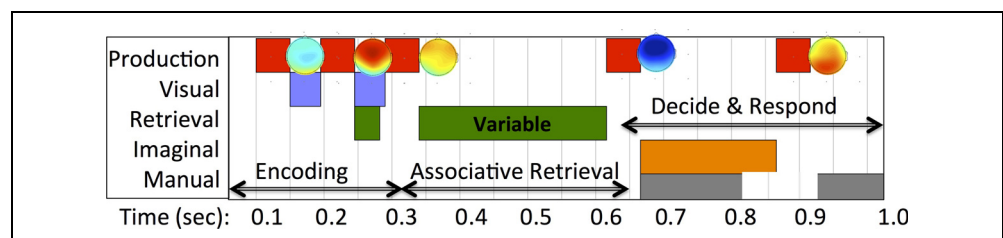
Our model of associative recognition places the effect of probe type in the retrieval stage. However, the results from other studies suggest that the effects of probe type may extend to the comparison stage in a different and conflicting way. The more similar the retrieved item is to the probe, the longer the comparison process will be. Studies of perceptual decision-making consistently show strong effects of foil similarity in response latency (for a review, see Farrell, 1985). Foils that share more features with targets are rejected more slowly. Likewise, King and Anderson (1976) found an effect of foil similarity on response latencies in associative recognition.

According to ACT-R's activation equations, overlapping foils will result in greater activation for the retrieved memory and, consequently, shorter retrieval times. The finding that responses were slower for overlapping foils indicates that the effect of probe similarity extends beyond the retrieval stage. We suspect that similarity also influences the comparison stage. Interestingly, and problematically for measures of overall RT, the similarity of the probe to a studied item in memory may reduce the duration of the retrieval stage and increase the duration of the comparison stage. If we can isolate these stages' durations using the HSMM-MVPA method, we can test two strong predictions about the effects of probe similarity:

1. Reflecting the effect of activation of the retrieved memory, the duration of the retrieval stage should decrease with the degree of match (i.e., probe similarity) between the probe and the studied pairs.
2. Reflecting the number of comparisons needed to detect a mismatch between the retrieved memory and the probe, the duration of the comparison stage should increase with the degree of match between the probe and the studied pairs.

To test these predictions, we conducted an experiment with a study phase and a test phase. In the study phase, participants learned 24 word triples. In half of the triples, both the person and location had one associate (Fan 1), and in half of the triples, both the person and

Figure 1. Swimlane representation of the ACT-R retrieve-to-reject model for targets and foils, with the scalp profiles identified from the HSMM-MVPA method.



location had two associates (Fan 2). All verbs had three associates. In the test phase, during which EEG was recorded, participants completed an associative recognition task where they distinguished between word triples they had studied (targets) and those they had not (dissimilar foils, similar 1 foils, and similar 2 foils). By using word triples instead of word pairs, we could examine the effects of multiple levels of probe similarity on the durations of the retrieval and comparison stages with the use of the HSMM-MVPA (Anderson et al., 2016).

METHODS

Participants

Twenty individuals from the Carnegie Mellon University community participated in a single 3-hr session for monetary compensation (14 men and 6 women, ages range from 19 to 35). All but one were right-handed. None reported a history of neurological impairment. One participant with less than 30% artifact-free EEG recording was excluded, leaving a total of 19 participants.

Materials

Participants memorized word triples composed of a person, a verb, and a location, shortened as “P-V-L” (e.g., *Musician-Walk-Factory*). The triples were created from lists of 18 people, 8 verbs, and 18 locations (Appendix D). Word length was four to eight letters for people, four to five letters for verbs, and four to seven letters for locations. Each participant memorized a list of 24 randomly generated word triples during the study phase of the experiment.

Foils were created by combining a person, verb, and location that had been studied in different triples from one another. Words were only swapped with other words that had the same number of associates to preserve the fan manipulation (e.g., Fan 2 foils were created using people and locations from other studied Fan 2 triples). There were three types of foils (Table 1): (1) dissimilar foils—none of the words in the triple had been studied together; (2) similar 1 foils—P-L had been studied together, but P-V and P-L had not; and (3) similar 2 foils—P-V and V-L had been studied together, but P-L had not. In total, participants were tested on 72 word

triples (24 targets, 24 dissimilar foils, 12 similar 1 foils, and 12 similar 2 foils). Half were Fan 1, and half were Fan 2 (Table 1). Foils were constructed such that every Fan 1 Person and Location appeared equally often, every Fan 2 Person and Location appeared equally often, and every Verb appeared equally often. Appendix II contains an example of 24 studies triples (targets) and a set of foils.

In addition to manipulating probe type and associative fan, we varied the vertical ordering of the words on the screen during the test phase. Half of the triples appeared in the studied order (P-V-L), and half appeared in a shuffled order (i.e., P-L-V, L-P-V, L-V-P, V-P-L, or V-L-P). Participants were instructed to decide whether the words had been studied together regardless of the ordering.

Procedure

The experiment began with a study phase where participants learned a list of 24 word triples. When a triple was presented for the first time, it appeared at the center of the screen for 8000 msec. Participants were instructed to read and memorize the triple. After all triples appeared once, participants advanced into the dropout portion of the study phase. In each trial, two of the three words in a triple appeared, and participants were required to recall and type the omitted word. There was no time limit to respond. If the response was incorrect, the correct answer was displayed for 4000 msec, and if the answer was correct, the word CORRECT appeared for 4000 msec. If a triple elicited an error, it was repeated again after all other triples had appeared. A block of trials ended when all triples had elicited a correct response. The dropout portion consisted of three blocks of trials. A different word from each triple was omitted during each block.

In the subsequent test phase, participants completed an associative recognition task where they distinguished between word triples they had studied and those they had not. Each trial began with a centrally presented fixation cross for a variable duration (sampled uniformly from 400 to 600 msec). Next, a probe word triple appeared vertically on the screen. Participants responded with a key press to indicate whether or not the triple had been studied. To respond “yes,” they pressed the J key with the right index finger, and to respond “no,” they pressed the K key with the right middle finger.

Table 1. Probe Types, Item Matches, and Probe Counts for Fan 1 and Fan 2 Triples

<i>Probe Type</i>	<i>Person-Verb Match?</i>	<i>Person-Location Match?</i>	<i>Verb-Location Match?</i>	<i>Fan 1</i>	<i>Fan 2</i>
Target	Yes	Yes	Yes	12	12
Dissimilar	No	No	No	12	12
Similar 1	No	Yes	No	6	6
Similar 2	Yes	No	Yes	6	6

Participants were instructed to respond as quickly and accurately as possible. This was reinforced with a bonus point system. One hundred points were deducted for every incorrect response, and $100-50 \times \text{Time}$ points were awarded for every correct response (if negative, the participant was given 0). *Time* was in seconds. After the participant responded, feedback appeared on the screen for 1000 msec. If the answer was incorrect, the word INCORRECT appeared along with the correct answer and the amount of points deducted. If the answer was correct, the word CORRECT appeared along with the amount of points awarded. Participants completed a total of 12 blocks of 72 trials. Each of the 72 word triples appeared once per test block.

EEG Recording and Analysis

Stimuli appeared on a 60-Hz LCD monitor set 60 cm from participants. The EEG was recorded from 128 Ag-AgCl sintered electrodes (10–20 system) using a Biosemi Active II System (BioSemi, Amsterdam, Netherlands). The EEG was re-referenced online to the combined common mode sense and driven right leg circuit. Electrodes were also placed on the right and left mastoids. Scalp recordings were algebraically re-referenced offline to the average of the right and left mastoids. The EEG and EOG signals were filtered with a bandpass of 0.1 to 70.0 Hz and were digitized at 512 Hz. The EEG recording was decomposed into independent components using the EEGLAB FastICA algorithm (Delorme & Makeig, 2004). Components associated with eye blinks were automatically identified and projected out of the EEG recording. Epochs of 1100 msec (including a 100-msec baseline) were then extracted from the continuous recording and corrected over the prestimulus interval. Epochs containing voltages above $+100 \mu\text{V}$ or below $-100 \mu\text{V}$ were excluded ($<4\%$ epochs).

EEG data were analyzed from trials with correct responses. Data were averaged across contiguous electrodes to create four regions: a left anterior/superior (LAS) region (FFC3h, F3, F1, FFC5h, FFC1h, FC3, and FC1), a right anterior/superior (RAS) region (FFC2h, F2, F4, FFC4h, FFC6h, FC2, and FC4), a left posterior/superior (LPS) region (CP3, CP1, CPP5h, CPP3h, CPP1h, P3, and P1), and a right posterior/superior (RPS) region (CP2, CP4, CPP2h, CPP4h, CPP6h, P2, and P4). We analyzed data from these regions during an early time window (300–500 msec) corresponding to the FN400 and a later time window (600–1000 msec) corresponding to the parietal old/new effect. Data from each time window were entered into a 4 (probe type) $\times 2$ (fan) $\times 2$ (laterality) $\times 2$ (anterior/superior) repeated-measures ANOVAs. For all analyses involving probe type (the only factor with more than two levels), we adjusted the p values using the Greenhouse–Geisser correction. We also analyzed response-locked waveforms, which were baseline-corrected using the 100-msec prestimulus interval before the triple appeared.

HSMM-MVPA Applied to EEG

In our HSMM, we explicitly model the variability of endogenous ERP components that would otherwise be distorted or lost in the average waveforms. The HSMM-MVPA method identifies brief, distinctive profiles of scalp activity (i.e., bumps) with variable latencies in the single-trial EEG (Anderson et al., 2016). A bump is modeled as a half-sine multidimensional peak across the scalp that signifies a significant change in the information processing, followed by a flat period where the mean of the ongoing sinusoidal noise is 0. Our HSMM models the durations of the flats as gamma distributions.

Two steps of dimensionality reduction were carried out to simplify the analysis and make the computations more efficient and tractable. First, the data were down-sampled to 100 Hz (i.e., 10-msec samples). Second, to deal with the highly intercorrelated nature of the EEG sensors and to reduce the dimensionality of the signal, spatial PCA (i.e., across electrodes) was performed to generate orthogonal PCA dimensions. The first 10 PCA components were retained. These accounted for 69.2% of the variance in the signal. The PCA components were z -scored for each trial. As a result, the data for the analysis consisted of 10 orthogonal PCA components sampled every 10 msec and with constant mean and variability across trials. Five samples (50 msec) beyond the response were also included in the analysis to ensure that the bump signifying the motor response was fully modeled, in the case that it occurred at the moment of trial completion. We only considered data from correct trials.¹

As described in more detail in our previous application of the HSMM-MVPA method (Anderson et al., 2016), several assumptions about the temporal structure of the signal are made to facilitate the analysis. First, the bumps were given a 50-msec width (i.e., five samples) with a half-sine shape. Such narrow bumps promote precision in the identification of stage boundaries, even if the bumps may actually be somewhat wider than 50 msec. Second, the analysis assumes that bumps do not overlap. Third, the flat durations are modeled as a gamma distribution with a fixed shape parameter 2 and a free scale parameter estimated to fit the data. See Anderson et al. (2016) for a detailed discussion of these assumptions, and tests of the robustness of the method against violations of each. Bumps in the HSMM are intended to account for the portion of the EEG signal corresponding to task-related processing; that is, variability arising from stimulus processing, memory retrieval and decision-making, and response commission. Other sources of variability in the EEG signal that are not accounted for by bumps in the HSMM include noise from muscle movement, ambient electrical activity in the recording environment, stochasticity in neural responses to the same or related events, and additional neural processes unrelated to the task that take place in a nonstimulus locked fashion.

An n bump HSMM requires estimating $n + 1$ stage distributions to describe the durations of the flats plus the n five-sample bumps for each PCA component. A different magnitude is estimated for each of the n bumps along each PCA dimension. A bump extends temporally across five samples (50 msec) and is multiplied by weights of 0.309, 0.809, 1.000, 0.809, and 0.309 (i.e., a 10-Hz half sin wave). The best model fit of such HSMMs is given by maximizing the summed log likelihood of the bumps and flats across all trials. For each trial, this log likelihood can be decomposed into two parts: the likelihood of the EEG data given that the bumps are centered at each time point, and the likelihoods that the bumps are centered at those time points given the gamma distributions that constrain their locations. In other words, the HSMM must select bump locations within a trial to maximize the correspondence between the observed and the estimated EEG signal, while selecting relatively consistent flat durations across trials to maximize their fit to the gamma distributions. The estimation process has to consider all possible combinations of bump locations, and this is what is efficiently calculated by the dynamic programming associated with HSMMs (Yu, 2010).

The HSMM methods also return the probabilities of each bump occurring at each time point on a trial-by-trial basis. These probabilities can be used to calculate the most likely location of each bump in a trial, which is the sum of the time points in the trial multiplied by the corresponding probability that the bump occurred at that time. Mean stage durations for a particular subject can then be calculated as the average time between bumps across all trials within that subject.

RESULTS

Behavioral Results

During the study phase, the number of times a triple was presented during each block before being correctly completed can be used to assess rate of learning (Table 2).² With a repeated-measures ANOVA with fan and block as factors, mean frequency decreased across blocks, $F(2, 36) = 30.382, p < .001$. The frequency was higher for Fan 2 triples versus Fan 1 triples, $F(1, 18) = 13.500, p < .001$. The overall effect of fan decreased slightly across block, with an interaction between block and fan, $F(2, 36) = 3.322, p < .05$. By the third block, participants tended to respond correctly on their first attempt.

Table 2. Mean Frequency of Triples with SEMs in Parenthesis

	<i>Block 1</i>	<i>Block 2</i>	<i>Block 3</i>
Fan 1	2.34 (0.23)	1.58 (0.11)	1.25 (0.07)
Fan 2	2.66 (0.25)	1.62 (0.09)	1.39 (0.09)

Table 3. Mean RT (in sec) and Error Rates (as a %) with SEMs in Parenthesis

<i>Probe Type</i>	<i>RT</i>		<i>Error Rate</i>	
	<i>Fan 1</i>	<i>Fan 2</i>	<i>Fan 1</i>	<i>Fan 2</i>
Target	1.26 (0.05)	1.42 (0.07)	13.3 (2.0)	27.9 (3.2)
Dissimilar	1.24 (0.06)	1.36 (0.07)	2.7 (0.1)	6.4 (1.2)
Similar 1	1.30 (0.06)	1.38 (0.08)	13.1 (2.7)	8.8 (1.4)
Similar 2	1.34 (0.06)	1.42 (0.07)	14.4 (2.2)	15.9 (1.8)

During the test phase, data were trimmed by excluding trials with RTs shorter than 40 msec or longer than 3 sec (4% of all trials). The mean correct RTs and error rates appear in Table 3. RT and error rate were submitted to a repeated-measures ANOVA with fan and probe type as factors. RT was longer and error rate was higher for Fan 2 triples versus Fan 1 triples, reflecting a main effect of fan on RT, $F(1, 18) = 7.066, p < .001$, and error rate, $F(1, 18) = 29.561, p < .001$. RT and error rate both varied with probe type ($F(3, 54) = 18.571, p < .001$ and $F(3, 54) = 11.965, p < .001$, respectively). Responses to dissimilar foils were fastest and most accurate. The effect of fan on performance was greatest for targets, reflecting an interaction between fan and probe type on RT, $F(3, 54) = 5.683, p < .01$, and error rate, $F(3, 54) = 28.654, p < .01$.

ERP Results

Stimulus Locked: 300–500 msec

Figure 2 shows ERPs from the four regions based on probe type and fan. We first analyzed the data from all four regions during 300–500 msec, the typical time window of the FN400 (Stróžak, Abedzadeh, & Curran, 2016; Mollison & Curran, 2012; Speer & Curran, 2007). Consistent with the impression conveyed by the figures, there were no significant main effects or interactions involving the experimental factors during this window (Table 4).

Stimulus Locked: 600–1000 msec

We then analyzed the data from all four regions during 600 to 1000 msec, the typical time window of the parietal old/new effect. Waveforms appeared more positive for targets, reflecting a main effect of probe (Table 4). The main effect of fan was not significant nor was the interaction between probe type and fan. The topographical distribution of the probe type effect is shown in Figure 3.

Response Locked: –50 to 50 msec

Effects not otherwise observable in stimulus-locked waveforms can sometimes be seen in response-locked

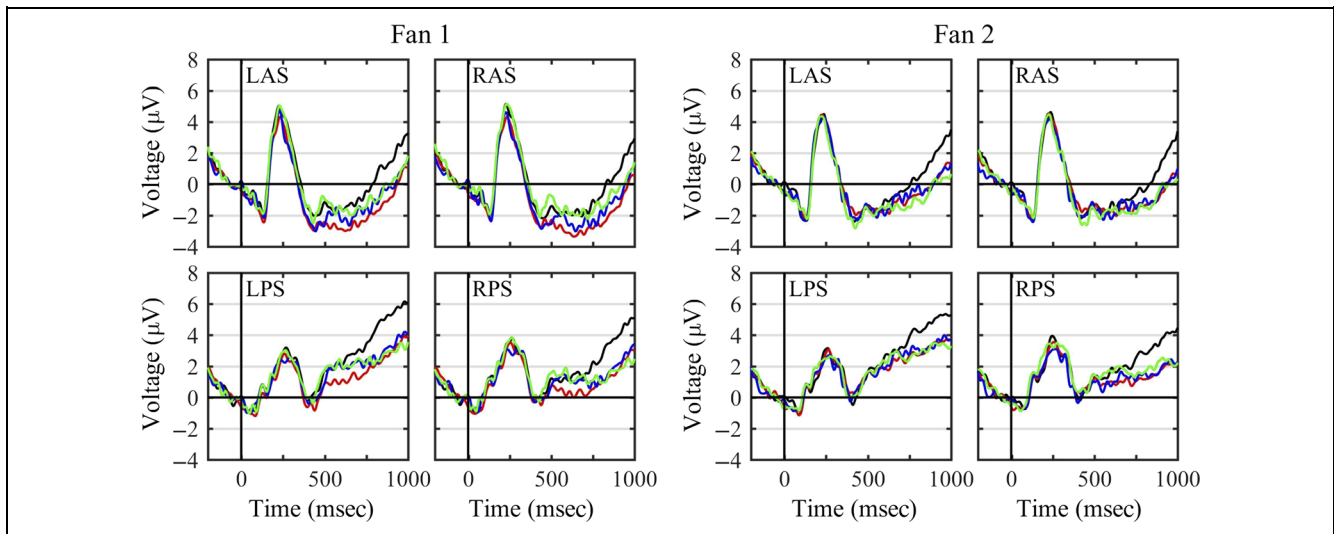


Figure 2. Stimulus locked waveforms from the four regions. Line colors correspond to target probes (black), dissimilar foils (red), similar 1 foils (blue), and similar 2 foils (green).

waveforms (Figure 4). A clear effect of probe type along with an effect of fan emerged surrounding the response from -50 to 50 msec. This was reflected in main effects of fan and probe (Table 4). The effect of fan was smallest for targets, reflecting an interaction between fan and probe. Waveforms were most positive for fan 1 triples and for targets. Figure 5 shows the topographical distribution.

In summary, we did not observe an FN400 in the stimulus-locked data. The time course, direction, and topographical distribution of the probe effect in the stimulus-locked data are consistent with the parietal old/new effect. The probe effect was also evident in the response-locked data, which is not surprising given that responses occurred shortly after the time window used to measure the parietal old/new effect. Finally, an effect of fan only appeared in the response-locked data. These results are consistent with our earlier study of paired associate recognition (Anderson et al., 2016). Now we

turn from the traditional analysis of averaging over trials to an HSMM-MVPA analysis that parses each trial into its stages.

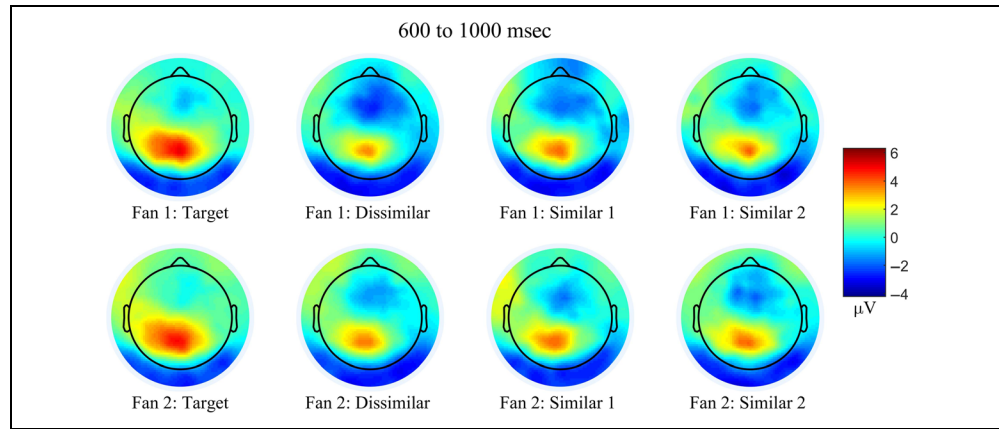
Identifying the Stage Durations and the Bump Profiles in HSMM-MVPA

HSMM-MVPA identifies bumps in the ongoing EEG signal related to significant changes in information processing. In this study, the number of stages in HSMM was decided jointly based on the EEG data and our theoretical understanding of the task model developed in ACT-R. Using LOOCV, we confirmed that an HSMM with four bumps outperformed any model with fewer bumps. When we included more than four bumps, the HSMM placed the additional bumps adjacent to one another to capture a sustained positivity in the EEG preceding the response, as also observed by Anderson et al. (2016). Given the absence of a theoretical motivation for including more

Table 4. 4 (Probe) by 2 (Fan) Repeated Measures ANOVAs for Mean Voltages during Stimulus and Response Locked Intervals

	Stimulus Locked 300 to 500 msec	Stimulus Locked 600 to 1000 msec	Response Locked -50 to 50 msec
Probe: $F(3, 54)$	0.715, <i>ns</i>	8.626, $p < .001$	5.826, $p < .01$
Fan: $F(1, 18)$	0.375, <i>ns</i>	2.348, <i>ns</i>	7.531, $p < .05$
Probe \times Fan: $F(3, 54)$	0.947, <i>ns</i>	0.757, <i>ns</i>	4.678, $p < .01$
Probe \times Anterior/Posterior: $F(3, 54)$	0.456, <i>ns</i>	0.159, <i>ns</i>	4.430, $p < .05$
Probe \times Laterality: $F(3, 54)$	1.361, <i>ns</i>	2.117, <i>ns</i>	3.862, $p < .05$
Fan \times Anterior/Posterior: $F(1, 18)$	1.452, <i>ns</i>	0.757, <i>ns</i>	1.554, <i>ns</i>
Fan \times Laterality: $F(1, 18)$	0.001, <i>ns</i>	0.179, <i>ns</i>	0.735, <i>ns</i>

Figure 3. Mean voltage over scalp from 600 to 1000 msec after stimulus onset.



stages in the computational model, we stuck with the four-bump model. Furthermore, given that the current task involved associative recognition, as did the task reported in Anderson et al. (2016), their identification of stages motivated us to adopt a similar model here.

On the basis of the ACT-R model of the task, we obtained a total of four bumps and five stages in processing: (1) Preattention stage: The time from stimulus onset to the first bump reflecting the time for the visual signal to reach the brain and be attended to; (2) Encoding stage: The time from the first to the second bump reflecting the time to encode the stimulus; (3) Retrieval stage: The time from the second to the third bump reflecting the time to retrieve a memory for comparison; (4) Comparison stage: The time from the third to the fourth bump reflecting the comparison of the memory with the probe; (5) Response stage: The time from the fourth bump to the response reflecting response execution.

Figure 6 illustrates the durations of the five processing stages and the scalp topologies of the four bumps iden-

tified using the HSMM-MVPA method. Each of the bumps is modeled as a 50-msec half-sine multidimensional peak that can be projected back to the scalp given the known PCA projection weights. The Preattention, Encoding, and Response stages were relatively brief compared with the Retrieval and Comparison stages. The bump profiles and durations are similar to those in Anderson et al. (2016), with the exception of the comparison stage, which is considerably longer here. The longer comparison stage reflects the greater number of words that participants needed to compare and the demanding similarity manipulation used in this experiment.

RTs varied by condition (Table 3). These differences in response latency must show up in the durations of some of the stages. To determine which stages were affected by the experiment manipulations, we fit HSMMs with different stage durations to each condition. That is, we estimated scale parameters for the gamma-2 distributions of each stage separately for the different conditions while constraining the bump magnitudes to be the same. To

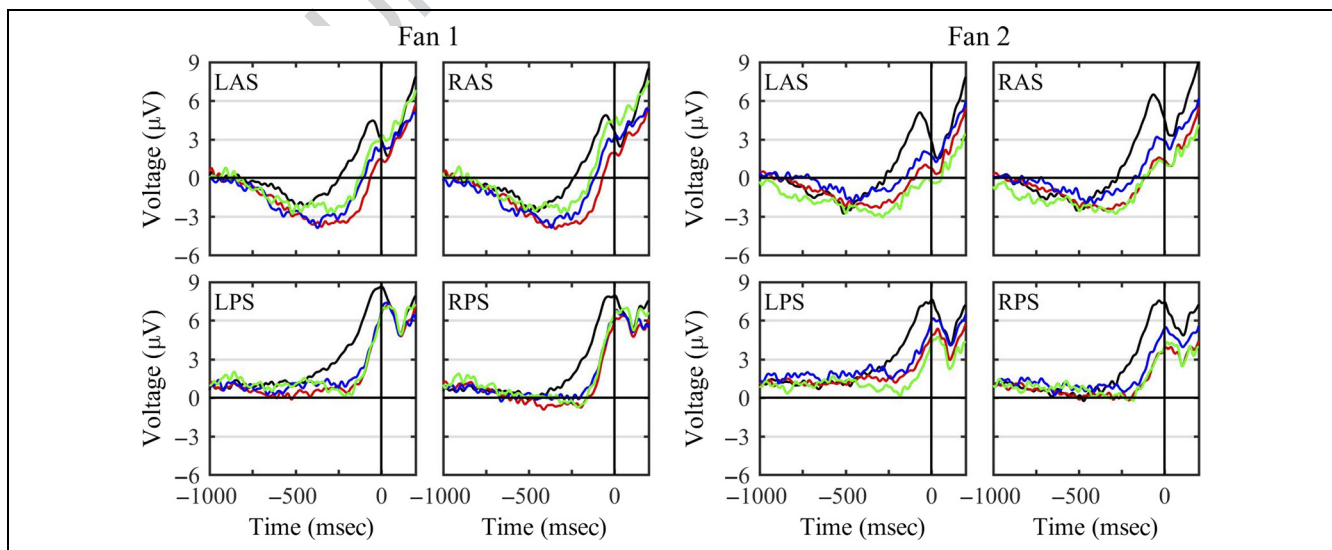
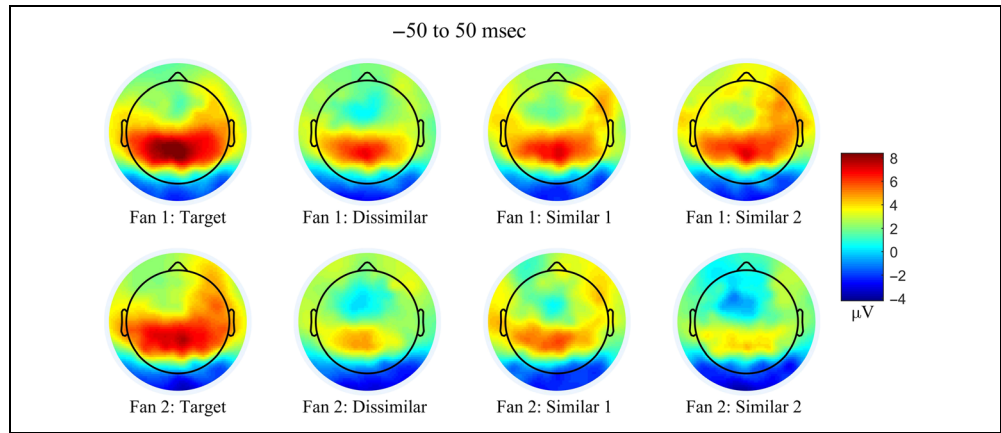


Figure 4. Response locked waveforms from the four regions. Line colors correspond to target probes (black), dissimilar foils (red), similar 1 foils (blue), and similar 2 foils (red).

Figure 5. Mean voltage over scalp from -50 to 50 msec with respect to response.



focus on the effects of probe type, we restricted the effect of fan to the Retrieval stage (Stage 3) and investigated which stages were impacted by probe type. To do so, we created an HSMM where the durations of all stages were allowed to vary with probe, but only the duration of the third stage was allowed to vary with fan. Figure 7 shows the resulting estimated stage durations. We computed the mean stage durations for each subject as a function of probe type and submitted them to a 4 (probe) repeated-measures ANOVA separately for each stage. Consistent with the impression conveyed by Figure 7, the effect of probe type was significant during the third stage, $F(3, 54) = 13.070$, $p < .0001$, and the fourth stage, $F(3, 54) = 21.720$, $p < .0001$. We subsequently performed a 4 (probe) \times 2 (stages three or four) repeated-measures ANOVA. The main effect of probe was significant, $F(3, 54) = 7.489$, $p < .001$, but the effect of stage was not, $F(1, 18) = 3.949$, $p < .1$. Most importantly, these two factors interacted, $F(3, 54) = 48.417$, $p < .0001$, owing to the opposing effects of probe on the durations of Stages 3 and 4.

The duration of Stage 3 was shortest for targets and comparable for the three foil types. This is consistent with the prediction of the ACT-R model that retrieval time will be the shortest for targets, because the retrieved triple always receives two sources of spreading activation (Equation 1 and Equation 2). However, the ACT-R model also predicts the slowest retrieval time for

dissimilar foils (which receive one source of spreading activation), followed by similar 1 foils and similar 2 foils. Probe similarity also affected Stage 4. In contrast to Stage 3, durations were longest for targets and similar 2 foils, followed by similar 1 foils, and finally by dissimilar foils. This is consistent with a model in which comparison terminates once a mismatch is detected.

Averaging across associative fan and probe type, RTs decreases from the first to the second half of the experiment (1374 vs. 1277 msec, $t(18) = 3.19$, $p < .01$). The ACT-R model predicts that practice should not affect the speed of encoding or motor planning, but rather the time to retrieve items from memory. Specifically, base level activation (B_i in Equation 2) increases with the number of repetitions. We compared the inferred stage durations from the first and second halves of the experiment in the HSMM. The results were partially in line with the model's predictions: the duration of the retrieval stage decreased from 527 to 498 msec, $t(18) = 4.994$, $p < .05$, but the duration of the comparison stage also decreased from 588 to 529 msec, $t(18) = 13.219$, $p < .01$.

Averaged Electrode Activity Anchored by Model Events

In standard ERP analyses, the EEG signal is anchored to observable events such as the presentation of a stimulus. The bumps obtained from the HSMM signal latent

Figure 6. Mean electrode activity reconstructed for the four bumps by projecting the bump magnitudes back to the PCA weights (top). Mean durations of five stages interleaved by four 50-msec bumps (bottom).

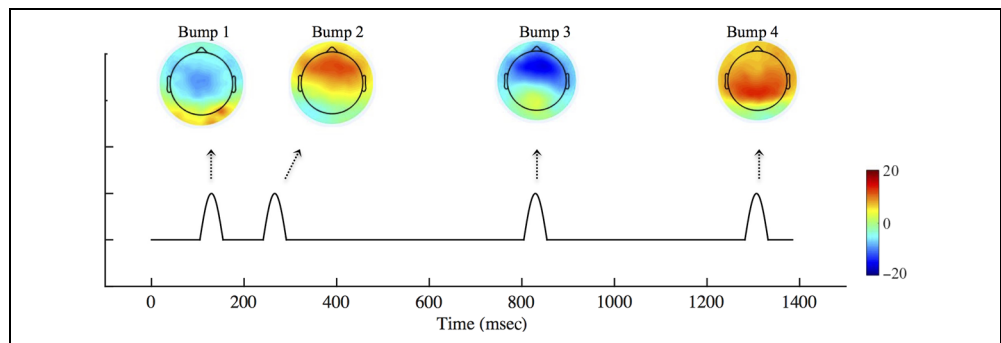
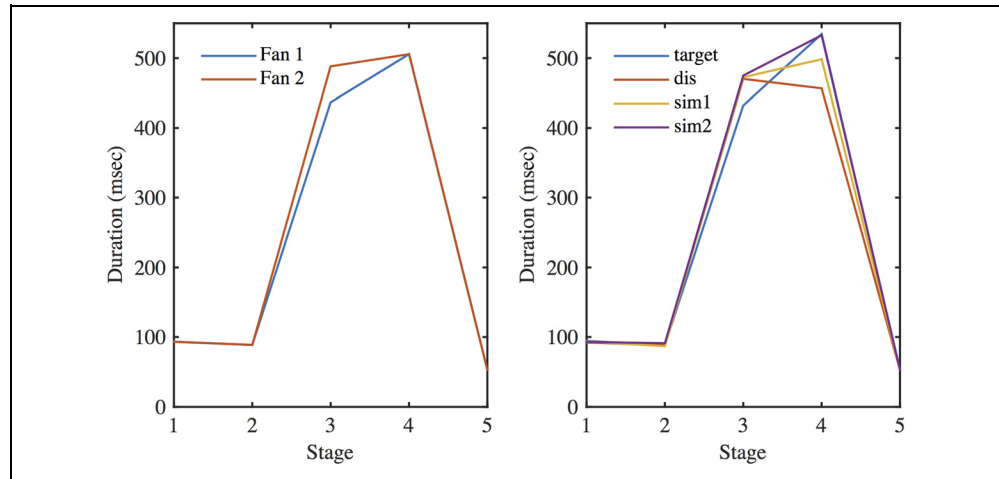


Figure 7. Condition-specific stage durations when localizing the effect of fan to Stage 3 and allowing probe type to vary across all stages.



points of change. These events can be used, along with observable events, to align the EEG data. We anchored the EEG data from each trial according to stimulus onset, response, and the maximum likelihood locations of each of the four bumps during that trial. We then expanded or contracted the resulting five intervals in every trial of a condition to have durations equal to those specified by HSMM. In this way, the stimulus, the locations of the four bumps, and the response are aligned across all trials in a condition. In the resulting displays (Figures 8 and 9), conditions with longer RTs stretch further forward in the stimulus-locked waveforms.

In Figures 8 and 9, substantial differences among conditions are seen before and during the fourth bump

over the parietal region. We performed a 4 (probe type) \times 2 (fan) \times 2 (flat/bump) repeated-measures ANOVA using mean amplitude over the flat preceding the fourth bump and peak amplitude of the fourth bump. There is a significant interaction of the three factors ($F(3, 54) = 6.072, p = .002$). Then we performed separately a 4 (probe type) \times 2 (fan) repeated-measures ANOVA using mean amplitude over the flat or peak amplitude of the fourth bump. The main effect of probe was significant both during the preceding flat ($F(3, 54) = 12.568, p < .001$) and during the fourth bump ($F(3, 54) = 6.593, p = .002$), reflecting the greater positivity for targets versus foils. The main effect of fan on the peak amplitude of the fourth bump was also significant ($F(1, 18) = 11.446, p = .003$),

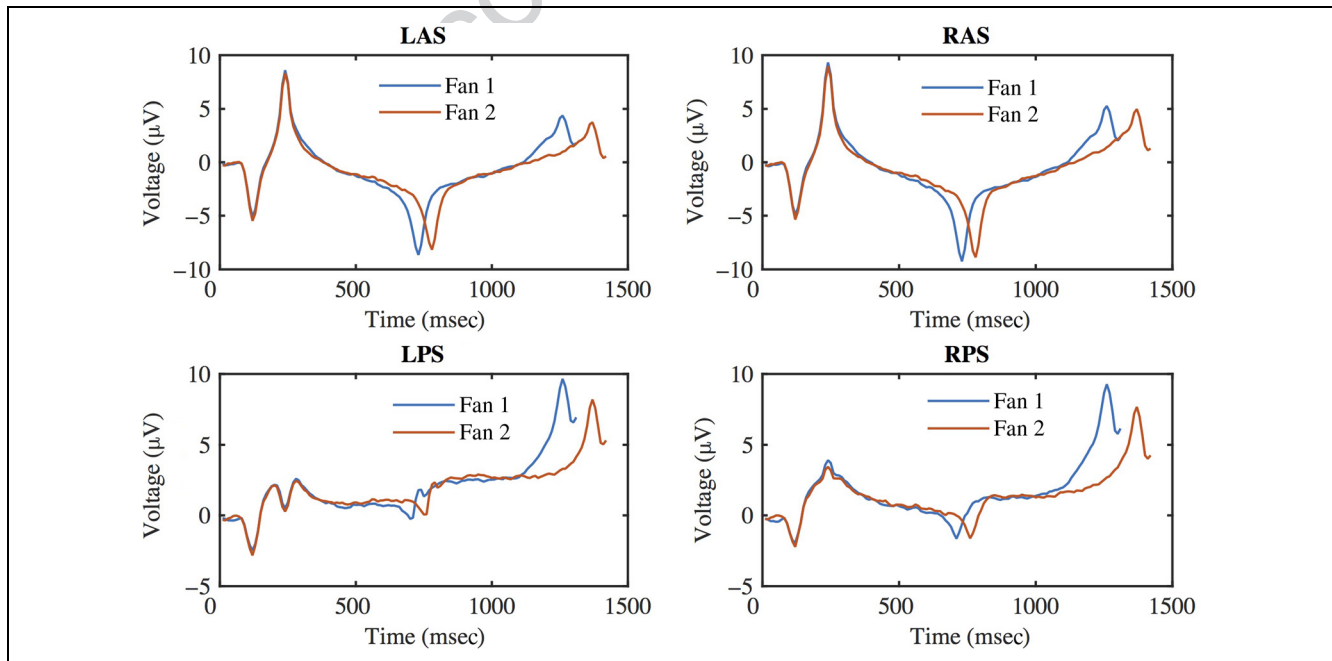


Figure 8. Average EEG data after warping every trial so that the maximum likelihood locations of the bumps correspond to the average locations for that condition. Data are shown in four regions (LAS, RPS, LPS, and RPS) for each fan condition.

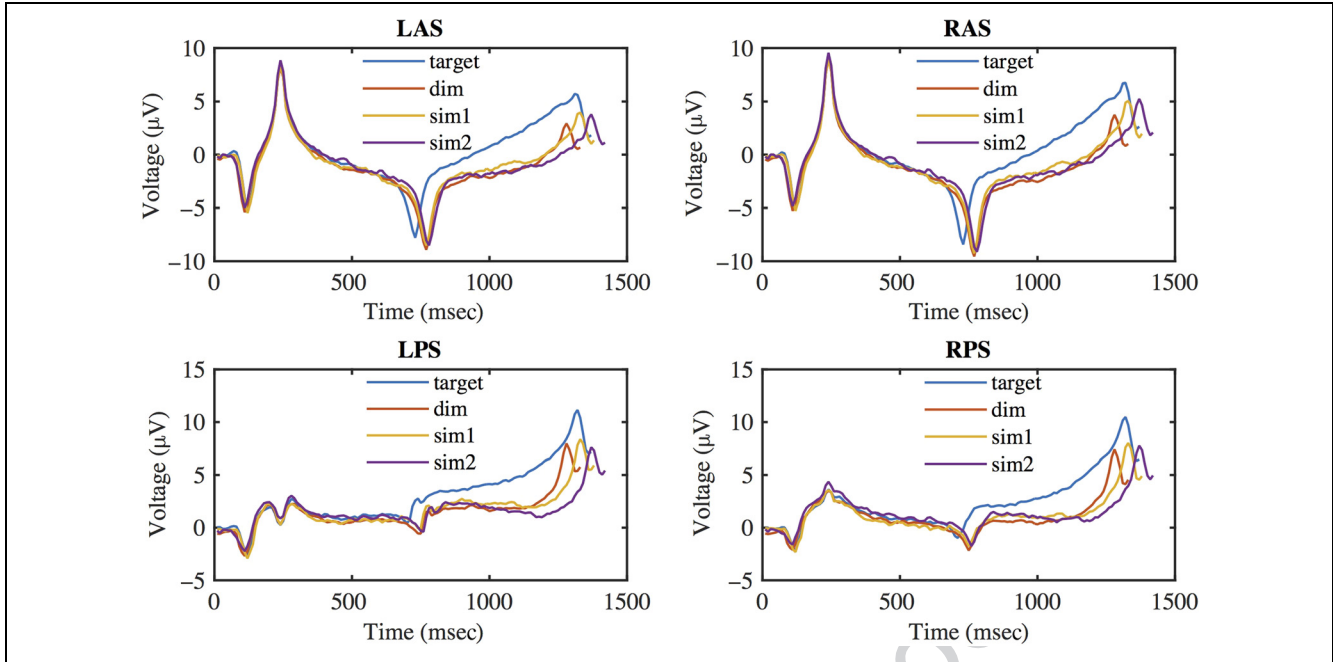


Figure 9. Average EEG data after warping every trial so that the maximum likelihood locations of the bumps correspond to the average locations for that condition. Data are shown in four regions (LAS, RPS, LPS, and RPS) for each probe condition.

owing to the greater positivity for Fan 1 triples versus Fan 2 triples.

Using the Neuroimaging Analysis to Inform the Task Model

Our model’s retrieve-to-reject strategy consists of four stages: (1) Encoding—randomly select two of the words to encode; (2) Retrieval—retrieve a triple from memory that most closely matches the two encoded words; (3) Comparison—compare the encoded and retrieved triples to determine whether any word differs; (4) Response—press K (“not studied”) if any word differs and press J (“studied”) otherwise. RTs equal the cumulative duration of the four stages. This can be represented as

$$RT_i = \text{Shared Time} + \text{Retrieval Time}_i + \text{Comparison Time}_i. \quad (4)$$

Shared Time includes the durations of the encoding and response stages. These encompass the time for the signal to reach the brain, the time to encode the words, and the time to program and perform a motor response.

Retrieval Time is computed from the set of equations that make up ACT-R’s theory of declarative memory and that are described in the Introduction.³ In this task, retrieval time depends on the number of cues (i.e., encoded words) that spread activation to the retrieved triple (Equation 1). The number of sources of activation in targets always equaled two because all possible pairs of encoded words (P-V, P-L, or V-L) occurred together in a studied triple. The number of sources of activation

for the retrieved triple in dissimilar foils always equaled one because no two of the words occurred together in a studied triple. The number of sources for similar 1 foils equaled 1 with 66% (P-V or V-L encoded) and 2 with 33% (P-L encoded). Lastly, the number of sources for similar 2 foils equaled 1 with 33% (P-L encoded) and 2 with 66% (P-V or V-L encoded).

Retrieval Time also depended on the number of associates or fan of the encoded words in the probe (Equation 3). Person and location always had the same fan as one another (1 or 2), which varied by triple. Verb always had a fan of 3. On the basis of the spread of activation (Equation 2) and the degree of match (i.e., probe similarity) between current context and the retrieved triple (Equation 1), we computed activation for all combinations of two encoded words for the four probe types and the two fan conditions. Retrieval time for a condition was calculated as the average expected retrieval duration for all combinations of two encoded words (P-V, P-L, or V-L encoded).

Comparison Time depended on the number of encoded words that matched words in the retrieved triple. We modeled the comparison stage as a serial process in which participant first compared the two encoded words to the corresponding words in the retrieved triple. If either word differed, the comparison process ended. If neither word differed, the participant compared the third word. The third word could conceivably be encoded before or during the final comparison.

The probability of comparing the third word varied by probe type. Targets always required comparing the third word because the first two words always matched

the retrieved triple. Dissimilar foils never required the final comparison because the first two words never both matched the retrieved triple. Similar 1 foils required the final comparison 33% of the time (P-L initially encoded). Lastly, similar 2 foils required the final comparison 66% of the time (P-V or V-L initially encoded). Time to compare the third word (*Final Comparison*) was treated as a free parameter. Overall comparison time for a condition was calculated as the sum of a *Comparison Intercept* parameter, which was the same across conditions and accounted for the duration of comparing the first two words, and the *Final Comparison* weighted according to the probability that the final word was compared in each condition.

Model Fitting Procedure

The ACT-R model contained a total of four free parameters: *Shared Time*, F (latency scale for mapping activation onto retrieval time), *Final Comparison*, and *Comparison Intercept*. Maximum associative strength (S , Equation 3) was set to a default value of 1.5,⁴ and attentional weight (W , Equation 2) was set to 1/2 because context was defined by the two encoded words.

We estimated the free model parameters to maximize the correspondence between the durations of the model stages (i.e., *Retrieval* and *Comparison*) and the stages inferred from the HSMM using a simplex optimization algorithm. This is in contrast to the standard approach to parameter estimation, which involves finding parameter values that maximize the correspondence between the model's output and the observed overall RTs. By isolating the durations of the retrieval and comparison stages using the neuroimaging data, it is possible to better estimate F (which only affects the retrieval stage) and *Final Comparison* (which only affects the comparison stage).

Model Results

The latency scalar parameter (F) was estimated to maximize the correspondence between the duration of the model's retrieval stage and the duration of the third stage in the HSMM-MVPA analysis. For the best fitting value of F (Table 5), the model accounted for the different durations of third stage across the eight conditions

Table 5. Model Parameter Estimates

Parameter Name	Estimate
<i>Shared Time</i>	430 msec
F	905 msec
<i>Final Comparison</i>	85 msec
<i>Comparison Intercept</i>	465 msec

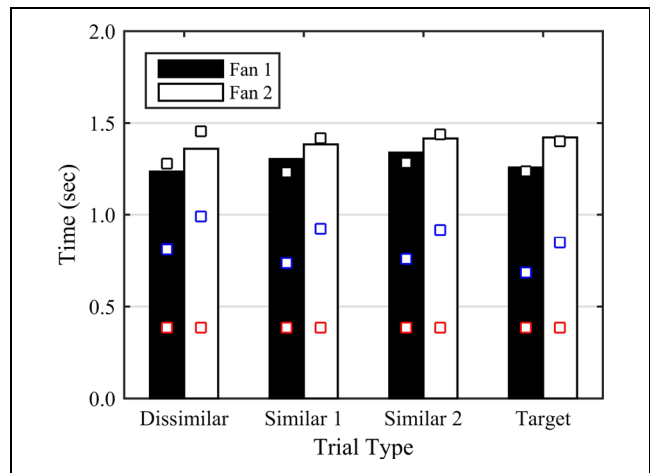


Figure 10. Observed RTs (bars) and cumulative duration of model's shared time (red squares), retrieval time (blue squares), and comparison time (black squares).

formed by fan and probe, as shown in Figure 7 ($r = .88$, $MSE = .005$).

The *Final Comparison* and *Comparison Intercept* parameters were estimated to maximize the correspondence between the duration of the model's comparison stage and the fourth stage in the HSMM-MVPA analysis. For the best fitting values (Table 5), the model accounted for the different durations of the fourth stage across the eight conditions formed by fan and probe, as shown in Figure 7 ($r = .94$, $MSE = .001$).

Using the parameter estimates from the neuroimaging data (Table 5), we calculated the three durations in Equation 4 (*Shared Time*, *Retrieval Time*, and *Comparison Time*) to generate model behavioral RTs. The correspondence between the model's overall RTs and the observed RTs was fairly high ($r = .81$, $MSE = .001$). As illustrated in Figure 10, fan only affected the duration of the model's retrieval stage. Alternatively, probe type affected the duration of the model's retrieval and comparison stages in opposite ways (Retrieval Stage: Target = 382 msec, Similar 2 = 453 msec, Similar 1 = 445 msec, Dissimilar = 515 msec; Comparison Stage: Target = 550 msec, Similar 2 = 522 msec, Similar 1 = 494 msec, Dissimilar = 465 msec). Owing to the strong negative correlation between the durations of the retrieval and comparison stages ($r^2 = -.86$), the net effect of probe type on overall RT was a small (Figure 10).

GENERAL DISCUSSION

We conducted a study of associative recognition in which participants decided whether a probe made up of three words matched any previously studied triple. We varied the associative fan of words in the probes, as well as the degree of similarity between probes and studied triples. Our HSMM-MVPA method revealed that probe

similarity substantially affected the durations of both the retrieval stage and the comparison stage. The retrieval stage was shortest when probes were more similar to a studied triple, whereas the comparison stage was longest when probes were more similar to a studied triple. The opposing ways in which probe similarity impacted retrieval and comparison stages explained why this factor had only a modest effect on overall RT; for instance, RTs were nearly identical for similar 1 foils and targets, yet the durations of the retrieval and comparison stages clearly differed.

A Model of Associative Recognition

The results from the HSMM-MVPA analysis were largely consistent with the ACT-R model of associative recognition. The model predicts that when a studied triple shares more words with a probe, the triple will be retrieved more quickly because of the greater number of sources spreading activation to it (Equation 2). Alternatively, the model predicts that when a probe has more words in common with a studied triple, serial comparison of the words in the probe and in the retrieved triple will take longer. This is because more words on average must be compared before detecting a difference. By isolating the durations of the retrieval and comparison stages using HSMM-MVPA, we were able to measure the effect of probe similarity on retrieval latency. Targets were retrieved more quickly than foils as predicted from the ACT-R model.

The results from the retrieval stage were not perfectly consistent with the model, however. The model predicts that dissimilar foils, which share only one word with any studied triple, will lead to slower retrievals than similar 1 or similar 2 foils, which share two words with some studied triple. In contrast, the HSMM-MVPA indicated that the duration of the retrieval stage was the same for all foils. This could reflect a simplifying assumption in our instantiation of the ACT-R model. Predictions based on Equation 2 correspond to the case of noiseless activation during retrieval. In its complete instantiation, the ACT-R cognitive architecture adds continuously varying, logistically distributed noise to activation values. Because many studied triples overlap with a dissimilar foil (six triples averaged over fan for dissimilar foils vs. one or two triples for the other foil types), the retrieved associate will be the most active of a larger set of candidates. As a consequence, the retrieved triple in the case of dissimilar foils will on average have a larger, positive noise term added to its activation value.

The HSMM-MVPA showed that the comparison stage was shortest for dissimilar foils and was longest for similar 2 foils and targets. This result relates to a classic finding from the “same”–“different” perceptual judgment task (Sternberg, 1969). The more attributes two visual stimuli share, the longer it takes for people to determine that they differ (Farell, 1985). This is consistent with a

serial comparison process, as instantiated in our ACT-R model of associative recognition. Interestingly, in the “same”–“different” task, matching stimuli produce faster responses than predicted by simple linear extrapolation of the number of comparisons required (Nickerson, 1969). As shown by the HSMM-MVPA, the duration of the comparison stage in our task increased with the number of comparisons for foils but was equivalent for targets and similar 2 foils even though targets required more comparisons. To account for the differential effects of number of comparisons on “same” versus “different” judgments, some models assume that “same” judgments are based on a parallel holistic process rather than a serial analytical process (Farell, 1985). The ACT-R model does not currently include a holistic comparison process and so predicts a slightly longer comparison stage for targets.

Comparing ERP Components and HSMM-MVPA Bumps of Associative Recognition

EEG studies of recognition memory reveal two ERP components, the FN400 and the parietal old/new effect. In many studies of recognition memory, participants view probes and are asked to decide whether they previously studied the probe. Some of the probes perfectly match a studied item (“targets”), some are entirely novel (“dissimilar lures”), and some merely resemble a studied item (“similar lures”). The FN400 is typically more negative for dissimilar lures than for targets or similar lures, suggesting that it is sensitive to item familiarity. The parietal old/new effect, on the other hand, is more positive for targets than for dissimilar or similar lures, indicating that it is sensitive to the retrieval of perfect matches (for a review, see Rugg & Curran, 2007).

Given that associative recognition involves remembering details about what an item appeared with, we expected that participants would display the standard parietal old/new effect. Indeed, as in previous studies of associative recognition (Diana, Van den Boom, Yonelinas, & Ranganath, 2011; Donaldson & Rugg, 1998), targets produced a late posterior positivity during the fourth flat that extended to the last bump, whereas foils did not. In the ACT-R model, the stage coinciding with the parietal old–new effect involves processing of the memory trace after its retrieval. We previously suggested that the amplitude of the sustained response reflected the different activations of the retrieved memories in the various conditions (Anderson et al., 2016). Activation is greater for targets than foils and for Fan 1 triples than for Fan 2 triples. Consistent with this view, voltages over the parietal scalp were greater for Fan 1 triples than for Fan 2 triples in the moments preceding a response.

All of the words in the test phase of our experiment appeared earlier during the study phase. As such, it was unclear whether participants would display an FN400 for foils versus targets. Some studies of associative recognition show that the novelty of the association between

probe items can produce an FN400 (Speer & Curran, 2007), whereas others do not (Anderson et al., 2016; Ecker, Zimmer, Groh-Bordin, & Mecklinger, 2007). These discrepancies may relate to whether the individual elements in an associate are represented as a single unit. Conditions that favor unitization in paired associate learning may yield a distinct FN400 for rearranged pairs relative to studied pairs (Ecker et al., 2007). These conditions include repeated study (Speer & Curran, 2007), semantically meaningful pairs (e.g., traffic-jam; Rhodes & Donaldson, 2007), and elaborative encoding (Rhodes & Donaldson, 2008). The conditions in our experiment did not promote unitization: Participants learned triples (rather than pairs), they studied triples as few as four times, and the words in triples were unrelated.

All regions show activity related to each of the four bumps. The topographical distribution and time course of the bumps are largely consistent with those described in our previous application of the HSMM-MVPA method to a paired associate recognition task (Anderson et al., 2016). The first bump likely corresponds to the N1, given its early time course, its anterior distribution, and its insensitivity to the fan and probe type manipulations. This component is typically interpreted as an index of visual attention (Luck, Woodman, & Vogel, 2000). The intermediate time course and anterior distribution of the second bump are consistent with the P2 (Van Petten, Kutas, Kluender, Mitchiner, & McIsaac, 1991).

The third bump may relate to the N2 (cf. Anderson et al., 2016), a frontocentral negativity caused by response conflict (Yeung, 2004). The N2 typically appears somewhat earlier in ERP waveforms. However, most studies of the N2 involve decisions far simpler than associative recognition. As the third bump in the ACT-R model initiates the comparison stage, it follows that this bump occurs at the moment of maximum response conflict within the trial. The late and variable latencies of the third bump may obscure the N2 in the conventional ERP waveforms of our study and in other studies of recognition memory. Finally, the time course, direction, and topographical distribution of the fourth bump are consistent with the parietal old/new effect. As in our previous experiment (Anderson et al., 2016), this bump was sensitive to perfect matches and had higher amplitude for Fan 1 triples than for Fan 2 triples.

The Path Forward: HSMM-MVPA

HSMM-MVPA can be used to guide the development of new theories by providing a direct measure of the durations of information processing stages to make inferences about the effects of experimental factors. RT-based methods have also been used. However, if an experimental factor affects the durations of multiple stages, its impact on each cannot be directly observed from overall RTs. Rather, one must specify a model of how the factor(s)

affect each stage, calculate the expected durations of all stages, and compare the summed stage durations to overall RTs. A discrepancy between expected and observed RTs does not indicate which stages and factors were modeled incorrectly. More problematically, the absence of a discrepancy does not exclude the possibility that the model overpredicted the duration of one stage and underpredicted the duration of another. By isolating each stage's duration, the HSMM method overcomes these limitations of RT-based methods.

HSMM-MVPA can be used for a second, related purpose: to obtain more accurate parameter estimates for a process model. An advantage of linking cognitive models to neural data would be the sheer wealth of additional information that neural data can provide in comparison with behavioral data (see a review of different linking approaches, de Hollander, Forstmann, & Brown, 2016). The HSMM-MVPA method provides individual stage durations instead of the sum of them (i.e., RT) to better constrain model fitting. For example, the ACT-R retrieval latency scalar (F) only affects the duration of the retrieval stage, whereas *Final Comparison* only affects the duration of the comparison stage. Because both parameters modulate the effects of probe type, changes in F can be partially offset by changes in *Final Comparison* time to produce similar overall RTs. By estimating model parameters based on the stage durations in the HSMM-MVPA, no such parameter compensation occurs. Another example is the estimation of *Comparison Intercept*, which is not affected by fan or probe type. Without information about the duration of the comparison stage from the HSMM-MVPA, it would not be possible to estimate the *Comparison Intercept* separately from other processes (e.g., encoding and responding) that are also not affected by the experimental factors. Beyond just a common intercept, Figure 7 illustrates that the method can break this time out into periods of prestimulus attention (Stage 1), encoding (Stage 2), and responding (Stage 5).

APPENDIX I

Table A1. People, Verbs, and Locations Used to Create Triples

	<i>Person</i> (18)		<i>Verb</i> (8)		<i>Location</i> (18)	
actor	coach	paint	airport	attic		
cowboy	dancer	work	bank	barn		
chef	engineer	walk	factory	garage		
farmer	musician	talk	hotel	kitchen		
maid	judge	sing	castle	library		
pilot	queen	laugh	museum	office		
lawyer	sheriff	sleep	church	prison		
soldier	teacher	drink	stadium	studio		
tourist	doctor	-	temple	theatre		

APPENDIX II

Table A2. List of 24 Studied Word Triples (Targets)

<i>Fan 1</i>	<i>Fan 2</i>
actor-paint-airport	musician-sing-kitchen
cowboy-paint-bank	judge-sing-library
chef-paint-factory	queen-sing-office
farmer-work-hotel	sheriff-laugh-kitchen
maid-work-castle	teacher-laugh-prison
pilot-work-museum	judge-laugh-studio
lawyer-walk-church	doctor-sleep-library
soldier-walk-stadium	queen-sleep-prison
tourist-walk-theater	sheriff-sleep-theater
coach-talk-attic	teacher-drink-office
dancer-talk-barn	doctor-drink-studio
engineer-talk-garage	musician-drink-theater

Table A3. Examples of Nonstudied Word Triples (Foils)

	<i>Fan 1</i>	<i>Fan 2</i>
Dissimilar	actor-work-church	musician-sleep-studio
Similar 1	actor-work-airport	musician-sleep-kitchen
Similar 2	actor-paint-bank	musician-sing-office

UNCITED REFERENCES

Bader, Mecklinger, Hoppstädter, & Meyer, 2010
 Sohn, Goode, Stenger, Carter, & Anderson, 2003

Acknowledgments

This research was supported by the National Science Foundation Grant 1420009, the James S. McDonnell Foundation Scholar Award 220020162, and the Office of Naval Research Grant N00014-15-1-2151 to J. R. A. and M. M. W.

Reprint requests should be sent to Qiong Zhang, Machine Learning Department, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213-3815, or via e-mail: qiongz@andrew.cmu.edu.

Notes

1. Though error trials are interesting in their own right, the current data set yielded too few to adequately constrain an HSMM-MVPA. Also, it is less clear what psychological events precipitated mistakes (e.g., retrieval errors or guessing).
2. More precisely, a pair of words was presented in each study trial, and participants retrieved the third word to complete the triple.
3. If all items are experienced equally often, B_j in Equation 2 can be set to zero. Furthermore, if equal attentional weight is assigned to each cue, W_j can be fixed to $1/n$ in Equation 2,

where n is the number of retrieval cues (i.e., two encoded words in the probe).

4. The selection of S is not important, because any change in S in Equation 3 will be offset during the estimation of F in Equation 1.

REFERENCES

- Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, *6*, 451–474.
- Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York: Oxford University Press.
- Anderson, J. R., & Fincham, J. M. (2014a). Discovering the sequential structure of thought. *Cognitive Science*, *38*, 322–352.
- Anderson, J. R., & Fincham, J. M. (2014b). Extending problem-solving procedures through reflection. *Cognitive Psychology*, *74*, 1–34.
- Anderson, J. R., & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General*, *128*, 186–197.
- Anderson, J. R., Zhang, Q., Borst, J. P., & Walsh, M. M. (2016). The discovery of processing stages: Extension of Sternberg's method. *Psychological Review*.
- Bader, R., Mecklinger, A., Hoppstädter, M., & Meyer, P. (2010). Recognition memory for one-trial-unitized word pairs: Evidence from event-related potentials. *Neuroimage*, *50*, 772–781.
- Borst, J. P., Schneider, D. W., Walsh, M. M., & Anderson, J. R. (2013). Stages of processing in associative recognition: Evidence from behavior, EEG, and classification. *Journal of Cognitive Neuroscience*, *25*, 2151–2166.
- Curran, T. (2000). Brain potentials of recollection and familiarity. *Memory & Cognition*, *28*, 923–938.
- Danker, J. F., Gunn, P., & Anderson, J. R. (2008). A rational account of memory predicts left prefrontal activation during controlled retrieval. *Cerebral Cortex*, *18*, 2674–2685.
- de Hollander, G., Forstmann, B. U., & Brown, S. D. (2016). Different ways of linking behavioral and neural data via computational cognitive models. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *1*, 101–109.
- Diana, R. A., Reder, L. M., Arndt, J., & Park, H. (2006). Models of recognition: A review of arguments in favor of a dual process account. *Psychonomic Bulletin & Review*, *13*, 1–21.
- Diana, R. A., Van den Boom, W., Yonelinas, A. P., & Ranganath, C. (2011). ERP correlates of source memory: Unitized source information increases familiarity-based retrieval. *Brain Research*, *1367*, 278–286.
- Donaldson, D. I., & Rugg, M. D. (1998). Recognition memory for new associations: Electrophysiological evidence for the role of recollection. *Neuropsychologia*, *36*, 377–395.
- Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, *30*, 412–431.
- Düzel, E., Yonelinas, A. P., Mangun, G. R., Heinze, H. J., & Tulving, E. (1997). Event-related potential correlates of two states of conscious awareness in memory. *Proceedings of the National Academy of Sciences, U.S.A.*, *94*, 5973–5978.
- Ecker, U. K., Zimmer, H. D., Groh-Bordin, C., & Mecklinger, A. (2007). Context effects on familiarity are familiarity effects of context—An electrophysiological study. *International Journal of Psychophysiology*, *64*, 146–156.
- Farell, B. (1985). Same-different judgments: A review of current controversies in perceptual comparisons. *Psychonomic Bulletin & Review*, *98*, 419–456.
- King, D. R. W., & Anderson, J. R. (1976). Long-term memory search: An intersecting activation process. *Journal of Verbal Learning and Verbal Behavior*, *15*, 587–605.

- Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.
- Luck, S. J., Woodman, G. F., & Vogel, E. K. (2000). Event-related potential studies of attention. *Trends in Cognitive Sciences*, 4, 432–440.
- Makeig, S., Westerfield, M., Jung, T. P., Enghoff, S., Townsend, J., Courchesne, E., et al. (2002). Dynamic brain sources of visual evoked responses. *Science*, 295, 690–694.
- Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology*, 57, 335–384.
- Nickerson, R. S. (1969). “Same”–“different” response times: A model and a preliminary test. In W. G. Koster (Ed.), *Attention and performance II*. Amsterdam: North-Holland.
- Pirolli, P. L., & Anderson, J. R. (1985). The role of practice in fact retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 11, 136–153.
- Rhodes, S. M., & Donaldson, D. I. (2007). Electrophysiological evidence for the influence of unitization on the processes engaged during episodic retrieval: Enhancing familiarity based remembering. *Neuropsychologia*, 45, 412–424.
- Rhodes, S. M., & Donaldson, D. I. (2008). Electrophysiological evidence for the effect of interactive imagery on episodic memory: Encouraging familiarity for non-unitized stimuli during associative recognition. *Neuroimage*, 39, 873–884.
- Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition*, 28, 907–922.
- Rotello, C. M., Macmillan, N. A., & Van Tassel, G. (2000). Recall-to-reject in recognition: Evidence from ROC curves. *Journal of Memory and Language*, 43, 67–88.
- Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, 11, 251–257.
- Schneider, D. W., & Anderson, J. R. (2012). Modeling fan effects on the time course of associative recognition. *Cognitive Psychology*, 64, 127–160.
- Sohn, M. H., Goode, A., Stenger, V. A., Carter, C. S., & Anderson, J. R. (2003). Competition and representation during memory retrieval: Roles of the prefrontal cortex and the posterior parietal cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, 100, 7412–7417.
- Sohn, M. H., Goode, A., Stenger, V. A., Jung, K. J., Carter, C. S., & Anderson, J. R. (2005). An information-processing model of three cortical regions: Evidence in episodic memory retrieval. *Neuroimage*, 25, 21–33.
- Speer, N. K., & Curran, T. (2007). ERP correlates of familiarity and recollection processes in visual associative recognition. *Brain Research*, 1174, 97–109.
- Stemberg, S. (1969). The discovery of processing stages: Extensions of Donders’ method. *Acta Psychologica*, 30, 276–315.
- Stróžak, P., Abedzadeh, D., & Curran, T. (2016). Separating the FN400 and N400 potentials across recognition memory experiments. *Brain Research*, 1635, 41–60.
- Van Petten, C., Kutas, M., Kluender, R., Mitchiner, M., & McIsaac, H. (1991). Fractionating the word repetition effect with event-related potentials. *Journal of Cognitive Neuroscience*, 3, 131–150.
- Walsh, M. M., Paynter, C. A., Zhang, Y., & Reder, L. M. (2016). Hitting the reset button: An ERP investigation of memory for temporal context. *Brain Research*, 1642, 524–531.
- Yeung, N., Bogacz, R., Holroyd, C. B., & Cohen, J. D. (2004). Detection of synchronized oscillations in the electroencephalogram: An evaluation of methods. *Psychophysiology*, 41, 822–832.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, 46, 441–517.
- Yu, S. Z. (2010). Hidden semi-Markov models. *Artificial Intelligence*, 174, 215–243.

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

During the preparation of your manuscript, the questions listed below arose. Kindly supply the necessary information.

1. Please check if section headings were correctly structured.
2. Danker (2010) was not cited in the reference list. Please check.
3. Please define LIPFC.
4. Delorme & Makeig, 2004, was not cited in the reference list. Please check.
5. “RESULTS” (section heading) was inserted here. Please check if appropriate.
6. Mollison & Curran, 2012, was not cited in the reference list. Please check.
7. Please define LOOCV.
8. Yeung, 2004, was not cited in the reference list. Please check.
9. Please insert citations of these references in the body: Bader, Mecklinger, Hoppstädter, & Meyer, 2010; Sohn, Goode, Stenger, Carter, & Anderson, 2003.
10. Please provide volume and page numbers of Anderson et al., 2016.
11. Please provide page number of Nickerson, 1969.

END OF ALL QUERIES