Assessing Student's Mental Representations of Complex Problem Spaces with EEG Technologies

Ronald H. Stevens, [1] Trysha L Galloway [1], Chris Berka [2], Robin Johnson [2] Marcia Sprang [3]

[1]IMMEX Project, UCLA School of Medicine, 5601 W. Slauson Ave., #255, Culver City, CA 90230
Ron@immex.com,
[2] Advanced Brain Monitoring, 2237 Faraday Avenue, Suite #100, Carlsbad, CA 92008
[3]Esperanza High School, 20683 Deodar Drive Yorba Linda, CA 92886

We have developed a neurophysiologic-based assessment of student's understanding of complex problem spaces that blends the population-based advantages of probabilistic performance modeling with the detection of neurophysiologic signals. It is designed to be rapid and effective in complex environments where assessment is often imprecise. Cohorts of novices, and experts encoded chemistry problem spaces by performing a series of online problem solving simulations. The stable memory encoding was verified by comparing their strategies with established probabilistic models of strategic performance. Then, we probed the neural correlates of the encoded problem space by measuring differential EEG signatures that were recorded in response to rapidly presented sequences of chemical reactions that represented different valid or invalid approaches for solving the chemistry problems. We found that experts completed performances in stacks more rapidly than did novices and they also correctly identified a higher percentage of reactions. Event related potentials revealed showed increased positivities in the 100-400 ms following presentation of the image preceding the decision when compared with the other stack images. This neural activity was used to explore reasons why students missed performances in the stack. One situation occurred when students appeared to have a lapse of attention. This was characterized by increased power in the 12-15 Hz range, a decrease in the ERP positivities at 100-400 ms after the final image presentation, and a slower reaction time. A second situation occurred when the students' decisions were almost entirely the reverse of what were expected. These responses were characterized by ERP morphologies similar to those of correct decisions suggesting the student had mistaken one set of chemical reactions for another.

INTRODUCTION

Encoding and consolidating problem spaces into memory require time and effort (Stevens et al, 2004, Soller & Stevens, 2007). A challenge facing researchers conducting neurophysiologic studies of subjects' encoding of problem spaces in self-paced environments is that there is considerable time between behaviorally relevant events (Gevins et al, 1997, Stevens et al, 2007). While navigating these problem spaces online there will be multiple mouse clicks, menu choices and even page changes as the problem solving event evolves. Such environments have a low neurophysiologic signal to noise ratio, complicating the deciphering of when information relevant to strategic problem solving is stably encoded into memory.

While associative learning and problem space encoding takes time, data retrieval during subsequent problem solving can be rapid. Broadly speaking, memory performance is a function of the degree to which cognitive operations engaged at encoding are recapitulated at retrieval (Tulving & Thomson, 1973, Nyberg et al, 2000), and there are anatomic (Reijmers et al (2007) and neurophysiologic (Rugg et al, 2000; Bastiaansen et al,, 2003) reasons to believe that retrieval of a student's representation of the problem space may activate similar pathways to those used to encode it. We believe that by studying the details of memory retrieval, valuable assessment information regarding the encoding process can be derived and used to create a closed loop feedback system for improving problem solving performance. In this manuscript we present initial efforts at developing such an assessment.

The assessment is being developed "on top of" IMMEX™, a library of online multimedia simulations for scientific problem solving that has a large user base (over 700,000 performances) and a refined set of probabilistic modeling tools for monitoring student's performance and progress (Cayetano et al, 2001, Stevens et al, 2004, Stevens & Soller, 2005, Soller & Stevens, 2007). The IMMEX™ (Integrated Multimedia Exercises) system presents case-study-type problems which students solve by searching multiple data and information sources. One sample task that we are using to develop the assessment is called *Hazmat*, which provides evidence of a student's ability to conduct qualitative chemical analyses (refs).

IMMEX™ and other simulation driven tasks continually involve decisions revolving around costs of obtaining information and deciding what specific items of information are needed next. These decisions drive a process of perceptual and diagnostic evaluation of stimuli and the integration of them into a running stream of potential outcomes. The most important decision during IMMEX™ problem solving, however, is the solution to the problem. Generally the decision is precipitated by a final piece of information that reduces the uncertainty to a decision threshold level.

This decision can be described in terms of (at least) two different contexts, the stimulus-locked processes linked to the ultimate piece of acquired information that reduces the uncertainty to a threshold level, and those locked to the decision itself (decision-locked).

Some decision-locked events will appear slightly before (several hundreds of milliseconds) the decision event such as the preparation for motor activity originating in the supplementary motor area. More distally, there appear to be cascades of decision events in the prefrontal and parietal cortex, which in the case of freely made intentions can precede decision awareness by up to 10s (Soon et al, 2008).

For stimuli, which in the case of IMMEX™ would be the display of a piece of information, the viewing of an image for as little as 27 ms. allows the detection and description of shades and shapes. By 100 ms an object can be recognized, and separation of objects and words can occur between 200 – 400 ms. Words then become recognized (around 500-600 ms), and later still (~750 ms) non-words are recognized.

This ability of the visual system to rapidly identify and classify objects has found practical applications for target recognition where subjects are shown rapid images which may or may not contain specific targets to be identified (Mathan, et al 2006). Power Event Related Potentials (PERP) and Event Related Potentials (ERP) collected from EEG sensors, particularly within the 3-7 Hz frequency range, can reliably determine whether or not the trainee recognized a target within an image (McKeeff & Tong, 2007).

Based on the above studies we believe that the rapid display of sequences of images that are linked by a problem solving context can leverage the speed of the human visual processing system to greatly accelerate our ability to assess student's representation of problem spaces.

METHODS

In our experiments, the subjects first encoded the *Hazmat* problem spaces by performing a series of IMMEX™ online chemistry problem solving simulations, often as part of normal classroom work. The stable memory encoding was verified by comparing their strategies with established probabilistic models of strategic performance (Stevens et al, 2004, Soller & Stevens, 2007). For memory retrieval, students were shown stacks in Rapid Sequence Visual Presentation (RSVP) mode (Gerson et al, 2005) of images from the *Hazmat* problem space which represented sequences of chemical reactions.

The student's task was to decide if the sequences represented the reactions for a particular compound. Each image presented was a screen shot from the *Hazmat* chemistry problem set centered in the display along with the text describing the test. A combination of *images* results in a sequence called a *performance* (containing 7-15 images), and the series of performances are arranged into *stacks* (containing 18-25 performances). In preliminary studies the image presentation rate was set to 600 ms, and subjects could vary the speeds from 300-3000ms. A 2 second blank screen separated each of the performances in the problem set and if the student did not make a decision by then, the performance repeats. The performances were all drawn from the IMMEX™ database which contains over 75,000 *Hazmat* performances.

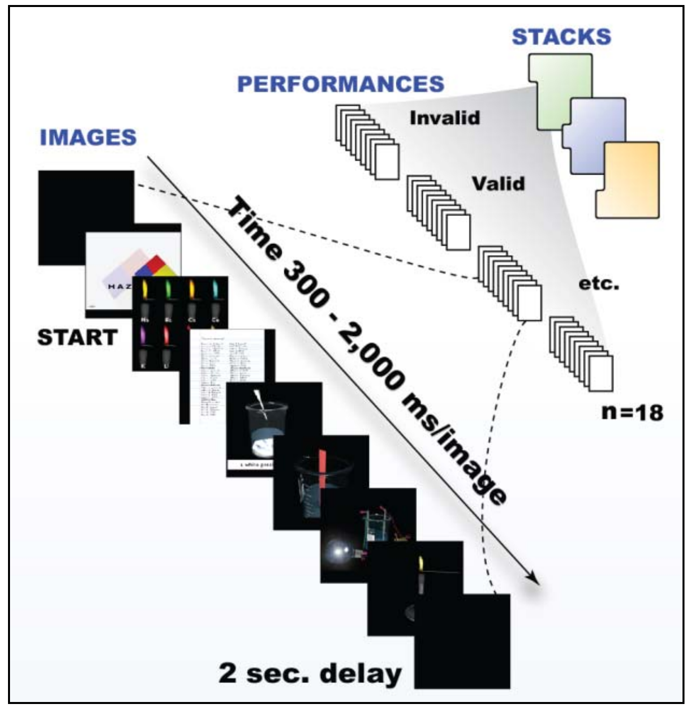


Figure 1 Experimental task for memory retrieval.

During the protocol, students login to the IMMEX™ server on the Internet and a Flash application is downloaded to the client machine along with the image stacks. When ready the stack sequences are presented. At the conclusion of the stacks (each containing 18-25 performance sequences), millisecond-level time stamped logs of each image presentations and student decisions were transferred to the IMMEX™ database server via Web Service technologies. This was used to merge the stack decision data with the Advanced Brain Monitoring (ABM) EEG files for neural signature analysis, and to relate student performance on the RSVP stacks with the performances and strategic approaches students used during the problem solving encoding phase.

Neural signatures of target detection, decision-making, saliency and accuracy of signal processing were obtained through the application of time-locked EEG potentials. Power spectral analysis and wavelet transformations were used to compute mean power spectra time-locked to a specific stimulus presentation or to a specific user response in the test bed environment. The EEG analysis window (between 250 and 5000msec.) was positioned to align with either a specific stimulus presentation event or over a response event to calculate the ERP associated with processing of the stimulus or with generation of the response. ABM software automatically labeled those single trial ERPs that were associated with significant artifacts including eye blinks, EMG, spikes, saturation or other contaminants and provides the option for including or excluding all trials with artifacts.

In addition to examining the single-trial and/or averaged ERPS as detected in the raw EEG signal, it is often useful to obtain the event-specific signatures using mean power spectra time-locked to the stimulus, response or other events of interest (termed "PERPs", Power Event-Related Potentials). To calculate the PERPs, the EEG is segmented into windows of at least 1000msec. The EEG analysis window is then positioned over either a specific stimulus presentation event or over a response event to calculate the EEG power associated with processing of a stimulus or with generation of a response. The EEG signal is also decomposed using a wavelets transformation, allowing the wavelets coefficients to also be used to characterize event-related EEG features.

RESULTS

Pilot Validation Studies

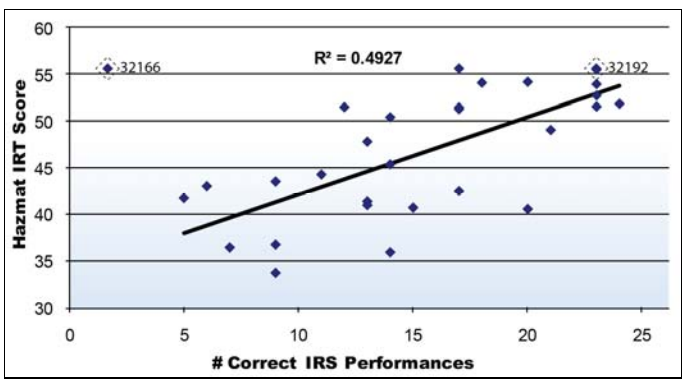


Figure 2. Correlation of Hazmat problem solving ability (item response theory estimates) and IRS performance.

An initial study sought to relate the IRS performance results to student's prior in-classroom problem solving performances The correlation between students' performance of the IRS tasks and their overall Hazmat problem solving ability was ($r^2 = 0.49$). This helped validate the IRS task as a reasonable approximation of the task used for problem space encoding. Further validation studies were conducted with five experienced / expert chemists (faculty – postdoctoral chemistry students) to establish the display speed of the images, and assess the ability of these chemists to distinguish valid sequences of chemical reactions from sequences of random images from *Hazmat* (i.e. invalid reactions). Each expert performed 3-5 stacks for a total of ~250 decision events. 95.3% of the decisions were confirmations or correct rejections. Most experts found the 600 ms image presentation speed too rapid and self-adjusted to a rate of 1200-1400 ms. As shown in Figure 3, with stack experience, the speed of distinguishing valid from invalid chemical sequences increased suggesting a training effect and valid performances were consistently identified slower than invalid performances. Subsequent studies with novices were conducted with the image presentation speed set at 1300 ms.

Studies comparing experts (n = 5) and novices (n = 75) showed that experts correctly identified more performances than novices (95% vs. 48%, p< 0.01), and completed performances more rapidly (19.8 sec. vs. 24.2 sec, p< 0.05).

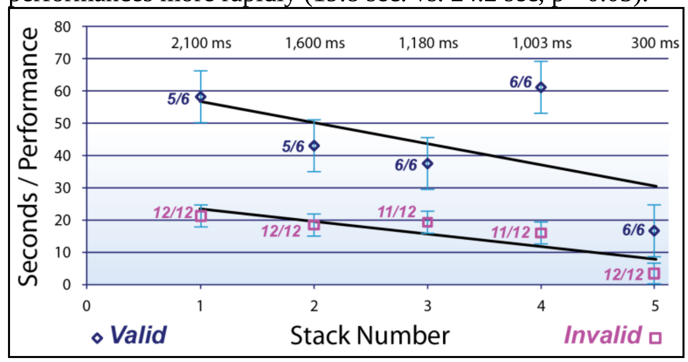


Figure 3 Stack Performance Metrics for Expert #700.

For most experts, the image in the stack immediately preceding the decision provided strong evidence for either confirming or refuting the evidence for the performance. From this a mean reaction time was constructed with an average of.92 ± .35 seconds (n = 14) for correct valid performances, 0.98 ± .43 (n = 32) for correct invalid performances and 1.05 ± .56 seconds (n = 8) for misses and/or false positives. These data suggest that close-range decision events can be studied by stimulus-locking to the presentation of the last image before the decision.

Identification of Neural Signatures at Decision Boundaries

We began by conducting an ERP analysis that was stimulus-locked to the last image. We call these 'threshold' images as it is at this point the uncertainty with the decision is reduced to the point where an overt decision is initiated. We then subtracted the contribution of all of the non decision images to obtain the difference. This resulted in increased positivity at 100 – 400 ms following the image display. There was also a large positive component around 700 ms after threshold image presentation (Figure 4).

The stimulus-locked ERPs were developed by locking the EEG signal to the epoch and data point of the last image before the decision. ERPs were then calculated (in micro volts) every 4 ms, for a total of 256 Hz sampling rate, for the 1 second following the presentation of the threshold image. The average ERP was also calculated for all the remaining images and this was subtracted from the values of the final image. The values are the averages from three student subjects who performed two stacks of 25 performances (150 decisions total).

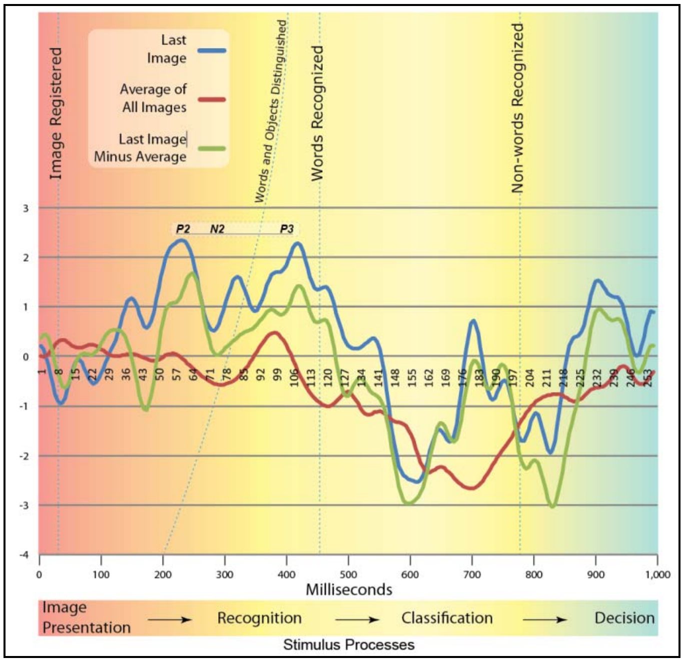
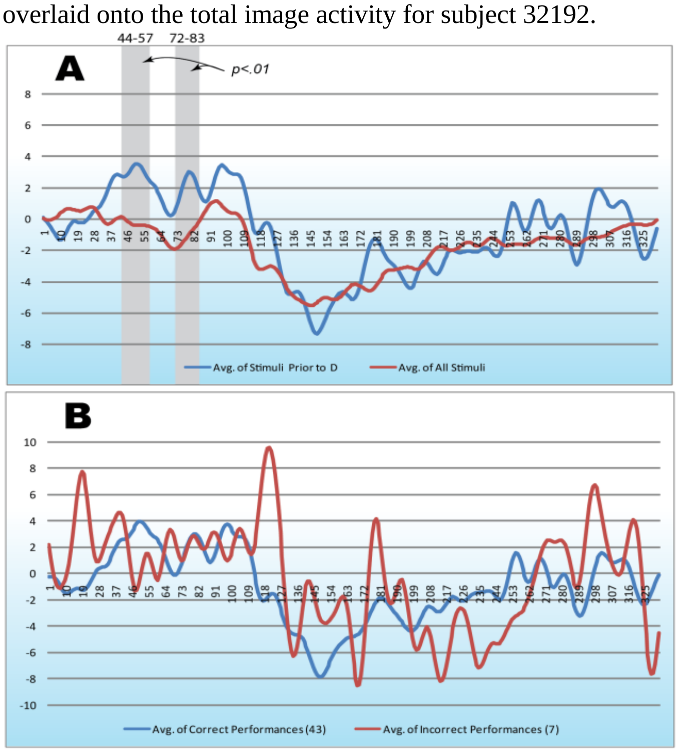


Figure 4. ERP for threshold images and all images in the stack.

Why Do Students Miss Problems?

We next examined the FzPO EEG activity for 1300 seconds following the threshold image, and separated these performances into the correct and incorrect responses. For this analysis we chose two students highlighted in Figure 3, both of whom solved the majority of the *Hazmat* problems (i.e. good encoding of the problem space), but showed large differences in their performance on IRS. Figure 5A shows the activity of the threshold images preceding a correct response overlaid onto the total image activity for subject 32192.



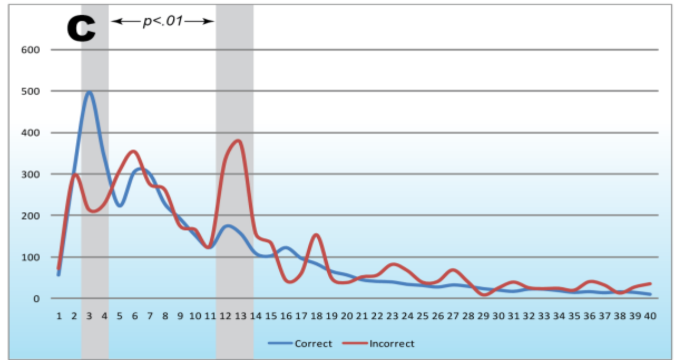


Figure 5. A and B. ERP C. PERP spectra for student 322192.

As expected, the correct responses (n =43) were preceded by significantly ($p< 0.01$) greater activity in data pts 44-57 and 72-83 (Figure 5A). In the incorrect responses (Figure 5B), (n = 7), this activity was replaced by a more regular 13-15 Hz rhythm. A power spectrum analysis of the correct and incorrect responses indicated increased activity in the 12-15 Hz range and a decreased activity in the 3-4 Hz range compared with correct responses. In addition, the reaction time was significantly longer for missed performances (2.3 ± 1.1 vs. 1.2 ± .99 sec. $p< 0.05$). Combined, these results suggest that the incorrect responses by this student may have resulted from momentary lapses in attention as described by Weissman et al (2006) and Eichel et al (2008).

The second student (#32166) consistently responded ‘No’ to all the performances that should have been ‘Yes’ and vice versa, apparently distinguishing the reactions correctly but confusing their meaning. As shown in Figure 6, the ERP activity of the threshold images in 31-93 was significantly more positive than that of all remaining images, suggesting that as she was choosing incorrectly, she believed she was making the correct decisions.

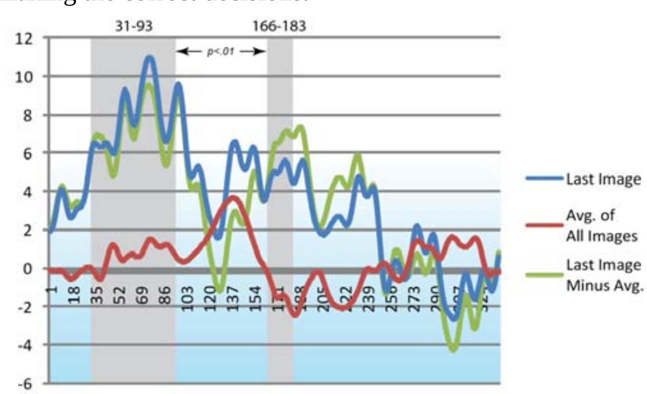


Figure 6. ERP for student 32166.

CONCLUSIONS

This study has presented an approach for studying how well students have encoded qualitative chemistry problem spaces and for exploring why students make incorrect decisions. The neural signatures detected helped identify at least two examples of reduced performance situations that may not have otherwise been observed. These situations, possible lapses of attention and correct use of incorrect content knowledge are likely to be common in novices.

The approach we have taken helps concentrate the number of student recognition and decision events that can be studied in a short period of time like a classroom environment, and helps improve the decision-related signal to noise ratio for a complex problem solving task . The correlation between the IRS performances and the *Hazmat* problem solving performances used for encoding suggests that the RSVP stacks may be a useful rapid approach for studying how well the problem space has been encoded, and for detecting possible neural signatures of the events leading up to a decision.

In this study we chose to lock the EEG to the last stimulus image before the decision event, rather than to the decision itself. This seemed justified as a) experts reject or confirm a performance around such key event images, b) the reaction time from these images to the decision is 1000 – 1200ms which is within the 1300ms display time for an image, and c) doing so allows the removal of most image-locked EEG activities that are not necessarily tied to the decision itself. Further studies are in progress to lock backward from decision itself which should give a different perspective of the events prior to the decision such as the preparation of motor activity for a key press.

We have shown that the major EEG difference between the threshold images and the other images was an increased positivity in the 100-400ms following the image display. This is when P2, N2 and P3 ERP components are observed following stimuli (Eichel et al, 2008). The relatively low resolution of the 6 channel bipolar headsets being used restricts our ability to resolve this activity further. Nevertheless, the activity over this time frame suggests the involvement of a matching process between the sensory input of the last image and the neuronal representation of stimuli in the context of the task. By monitoring this activity in a closed loop manner after the presentation of each image, we may be able to detect the onset of a decision, and intervene when the decision may be inappropriate.

REFERENCES

Bastiaansen, M. C. M., & Hagoort, P. (2003). *Event-induced theta responses as a window on the dynamics of memory*. Cortex, 39, 967-992.

Eichele, Tom., Debener, Stefan., Calhoun, Vince D., Specht., Engel, Andreas K., Hugdahl, Kenneth., Von Cramon, Yves D., Ullsperger, Markus. (2008). *Prediction of Human Errors by Maladaptive Changes in Event-Related Brain Networks.* Proc. Natl. Acad. Sci. USA Apr 21, 2008.

Eichele, Tom., Specht, Karsten., Moosmann, Matthias., Jongsma, Marijite L.A., Quiroga, Quian,. Nordby, Heige., Hugdahl, Kenneth. *Assessing the Spatiotemporal Evolution of Neuronal Activation with Single-Trial Event-Related Potentials and Functional MRI*. Proc Natl Acad Sci U S A. 2005 Dec 6;102(49):17798-803. Epub 2005 Nov 28.

Gerson, A., Parra, L.C., and Sajda, P. (2005). *Cortical origins of response time variability during rapid discrimination of visual objects.* Neuroimage, 28: 342-353.

Gevins, A., Smith, M. E., McEvoy, L., & Yu, D. (1997). *High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice*. Cerebral Cortex , 7, 374-385.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., Perrig, Wlater J., (2008). *Improving Fluid Intelligence with Training on Working Memory*. Proc. Natl. Acad. Sci. USA Apr 28, 2008; doi: 10.1073/pnas.0801268105.

Mathan, S., Ververs, P., Dorneich, M., Whitlow, S., Carciofini, J., Erdogmas, D., Pavel, M., Huang, C., Lan, T., and Adami, A. *Neurotechnology for Image Analysis: Searching for Needles in Haystack Efficiently.* In D.D. Schmorrow, K. Stanney (Eds): Augmented Cognition, v. 2, pp. 1-9, 2006.

McKeeff, T.J., and Tong, F. (2007). *The timing of perceptual decisions for ambiguous face stimuli in the human ventral visual cortex.* Cerebral Cortex, 17: 669-678.

Nyberg, l, Habib, R., McIntosh, A.R., & Tulving, E. (2000). *Reactivation of encoding-related brain activity during memory retrieval.* Proceedings of the National Academy of Science, USA. 97: 11120-11124.

Reijmers, L,G., Perking, B.L., Matsuo, N., and Mayford, M. (2007). *Localization of a stable neural correlate of associative memory*. Science 317: 1230-1234

Rugg, M.D., & Wilding, E.L. (2000). *Retrieval processing and episodic memory*. Trends in Cognitive Sciences, 4: 108-115.

Soller, A., and Stevens, R. H. (2007). *Applications of Stochastic Analyses for Collaborative Learning and Cognitive Assessmen*t. In Advances in Latent Variable Mixture Models, Gregory Hancock and Karen Samuelson (Eds.). Information Age Publishing.

Soon, C, S., Brass, M., Heinze, H., and Dylan Haynes, J. (2008). *Unconscious determinants of free decision in the human brain.* Nat. Neurosci. advanced online publication, 5 (13 Apr 2008).

Stevens, R. H., Soller, A., Cooper, M., and Sprang, M. (2004) *Modeling the Development of Problem Solving Skills in Chemistry with a Web-Based Tutor.* Intelligent Tutoring Systems. J. C. Lester, R. M. Vicari, & F. Paraguaca (Eds). Springer-Verlag Berlin Heidelberg, Germany. 7th International Conference Proceedings (pp. 580-591).

Stevens, R. H., Galloway, T., and Berka, C. (2007). *Exploring Neural Trajectories of Scientific Problem Solving Skill Acquisition.* Lecture Notes in Computer Science. Foundations of Augmented Cognition. Springer Berlin / Heidelberg. Volume 4565/2007, pp 400-408.

Tulving, E., & Thomson, D. M. (1973). *Encoding specificity and retrieval processes in episodic memory.* Psychological Review, 352-373.

Walther, D., Rutishauser, U., Koch, C., & Perona, P. (2005). *Selective visual attention enables learning and recognition of multiple objects in cluttered scenes.* Computer Vision and Image Understanding, 100 (1-2), 41-63 .

D H Weissman[1], K C Roberts[1], K M Visscher[2] & M G Woldorff[1] *The Neural Bases of Momentary Lapses in Attention*. , Nat. Neurosci. 9, 7 (Jul 2006).