Decoding Information Processing When Attention Fails: An Electrophysiological Approach

Ryan Kasper, Koel Das, Miguel P. Eckstein, Barry Giesbrecht

Department of Psychology
Institute for Collaborative Biotechnologies
University of California, Santa Barbara

ABSTRACT

The success of the attentional system in keeping people “on task” in dynamic environments arises from the coordinated operation of multiple neural networks. performance can occur. Here, we investigated the neural bases of attentional failuAlthough this coordinated effort is often successful, errors in res using computational techniques combined with high temporal resolution measures of brain activity using EEG. Attentional failures were induced by presenting two masked targets in rapid succession. In this task, correct identification of the first (T1) leads to impaired identification of the second (T2), a performance failure known as the attentional blink (AB). We applied linear pattern classification algorithms to measures of neural activity acquired during the AB to investigate two key issues about the temporal dynamics of visual attention. First, we tested whether the computational approaches would accurately discriminate the stimulus presented to the observer independent of behavior. Second, we tested whether our computational approaches could predict when the observer would make an error. Our analyses revealed that single-trial EEG activity could be used to not only predict the type of stimulus presented to the observer, but also to predict performance errors. These results are consistent with the notion that the brain represents information about the type of stimuli presented to observers and suggest that computational approaches may be used to provide a moment-by-moment analysis of an observer's attentional state.

Keywords: Attention, Pattern Classification, EEG, ERP, Augmented Cognition

INTRODUCTION

Whether you are a motorist driving on a busy street, an air traffic controller monitoring traffic at an airport, or a high-school student in an algebra class, the ability to selectively maintain one's attentional focus on task-relevant information while ignoring distracting information is vital for good performance. Although effective selective attention helps to keep us on task, errors in performance can sometimes occur because the capacity of the attentional system is limited. While attentional limitations are common in a variety of daily settings, they can be exacerbated by many factors, including learning disabilities, brain pathology or trauma, stress, task context, and individual differences. Thus, understanding the cognitive and neural mechanisms of these attentional limitations will facilitate their amelioration in the clinic, classroom, and the workplace.

The aim of the present work was to investigate the neural mechanisms of attentional failures by combining measures of neural activity and performance acquired during a difficult attention task with computational approaches that allow one to classify patterns of neural activity associated with different stimuli and cognitive states. To address this aim, we focused on one well-studied example of a limitation of the attentional system observed in the lab, known as the attentional blink (AB, Raymond, Shapiro, & Arnell, 1992). The AB is typically observed when two masked targets are presented in a rapid visual sequence. When the first target (T1) is identified, the identification of the second target (T2) is hindered for about 500 ms. There are two key characteristics of the AB that make it a powerful experimental tool for investigating the neural mechanisms of attentional limitations. First, the AB appears to require generalized attentional mechanisms that are capacity (or resource) limited. Consistent with the view that the AB involves generalized attentional systems, neuropsychological and neuroimaging studies have reported that the right hemisphere, which plays a large role in attentional control (e.g., Giesbrecht & Mangun, 2005; Giesbrecht, Woldorff, Song, & Mangun, 2003; Hopfinger, Buonocore, & Mangun, 2000) is also critically involved in the AB (e.g., Giesbrecht & Kingstone, 2004; Marois, Chun, & Gore, 2000; Marois, Yi, & Chun, 2004). Second, EEG studies have demonstrated that a fast, transient, yet robust temporal profile of the AB that provides moment-by-moment behavioral estimate of attentional demands emerges out of a complex pattern of neural dynamics that can be measured using multiple features of the EEG signal (amplitude, power, and phase; e.g., Slagter et al., 2007; Vogel & Machizawa, 2004).

The second component of our approach is the application of machine learning algorithms to measures of brain activity acquired during an AB task. Traditional behavioral and neuroimaging approaches are univariate, such that they typically use a single electrode or the average of a few electrodes rather than combining information neural information across electrodes. Combining neural information across electrodes could potentially provide key information about perceptual, attentional, and high-order cognitive states because they are more likely to be represented in terms of patterns of neural responses that may be best characterized in a multivariate data space. Computer classification algorithms are designed to extract information about patterns that discriminate between classes of information. These algorithms have been applied to both fMRI data and to EEG data to correctly identify the type of visual object shown to the observer (e.g., Haynes & Rees, 2005; Kamitani & Tong, 2005; Philiastides & Sajda, 2006a, 2006b).

We used the AB and pattern classification approaches to investigate the neural mechanisms of attentional failures in the following manner. First, subjects performed an AB task in which a context word was presented prior to a rapid serial visual presentation containing a T2 word that was either related or unrelated to the initial context word. The manipulation of context allowed us to focus on the N400 event-related potential (ERP) component, which has been previously shown to survive the AB despite the behavioral impairment (Giesbrecht, Sy, & Elliott, 2007; Luck, Vogel, & Shapiro, 1996; Rolke, Heil, Streb, & Henninghausen, 2001). Second, we used the ERP amplitudes recorded during this task as inputs into a linear pattern classifier to investigate two questions. First, we predicted that even though performance on the T2 is impaired, the pattern classifier should be able to discriminate the type of stimulus presented to the observer. Second, we predicted that to the extent that performance is represented in patterns of EEG responses, the pattern classifier should discriminate between trials on which the observer was correct versus when they were incorrect.

METHOD

Participants

Thirteen undergraduates from the University of California Santa Barbara were granted course credit or paid $10 per hour.

Procedure

Trials began with the presentation of a context word for 1000 ms, followed by a 750-1250 ms random delay, and then the RSVP stream. The stream consisted of a series of randomized character strings of uppercase black letters, each seven items long (~.8° x 2.5°). Each string was presented for 106 ms with no ISI. Within this stream there were two targets that were presented. T1 was a 7-item number string, all of which were the same parity, while T2 was a red word. T2 words that were not the full seven digits were flanked on either side by the letter X (e.g. XXHATXX). The temporal separation between T1 and T2, or lag, was either 320 ms or 960 ms. After the stream, there was a 750-1250 ms delay, followed by two response probes. The first prompted participants to report whether T1 consisted of odd or even numbers. The second prompted participants to indicate whether T2 was related or unrelated to the initial context word. Responses were untimed and made with a computer mouse. After response, fixation appeared again until the subject initiated the next trial. A schematic trial sequence is shown in Figure 1.

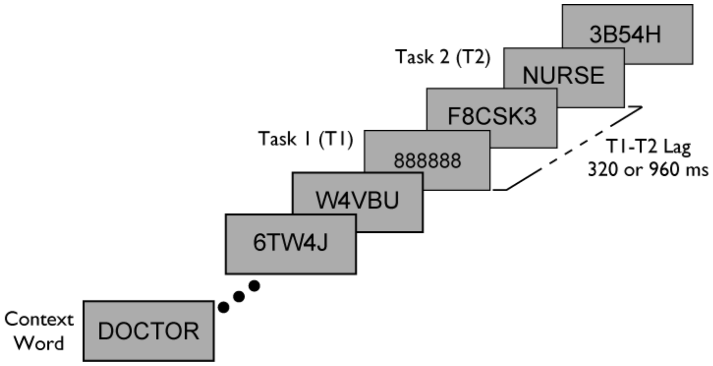


Figure 1. A schematic representation of the sequence of each trial.

Half of the trials contained semantically related context and T2 word pairs, while the other half were unrelated. The words used and the construction of the related and unrelated lists have been used in previous studies (e.g., Giesbrecht et al., 2007).

EEG Recording & Analysis

Recording of EEG was done at 256 Hz from 64 electrodes mounted in an elastic cap and positioned according to the 10/20 system. Electrodes were also placed above and below each eye for the vertical electrooculogram (EOG), as well as 1 cm lateral to the external canthi on each side for the horizontal EOG. The data were re-referenced offline to the average of the signal recorded at the left and right mastoids and then band-pass filtered (.01-100 Hz). The average ERP waveforms in all conditions were computed time-locked to T1 and T2 stimulus onset and included a 200 ms pre-stimulus baseline and 600 ms post-stimulus interval. Trials containing EOG artifacts from eye movements or blinks (±100μV) were rejected from further analysis.

The traditional ERP analyses involved averaging the segmented epochs for each individual in and in each condition. Following previous studies (Giesbrecht et al., 2007), we computed a difference wave that subtracted the response on related trials from the response on unrelated trials. Because the sensory stimulation in each case was exactly the same (only the context word differed), the difference wave reveals the effect of context, uncontaminated by the sensory response. Hypothesis tests were then conducted on the difference waves using a repeated measures ANOVA. The Greenhouse-Geisser correction for the degrees of freedom was used where appropriate. The pattern classification analyses used a standard linear discriminant analysis (LDA, e.g., Fisher, 1936) that computes the best fit linear weights for a set of training trials and then uses these weights to compute a weighted average across the input features for an independent set of test trials. A decision rule was then applied to the result to classify the input patterns. The inputs to the classifier were single trial responses at all 64 electrodes. To reduce the dimensionality, we averaged the single trial responses at each electrode into non-overlapping 20 ms time bins, starting with stimulus onset. The categories to be classified were T2 stimulus type (related vs. unrelated) and trial accuracy (correct vs. incorrect). In each case, the training data consisted of all but 20 of the experimental trials, which were set aside for testing. The performance of the classifier was evaluated using a 10-fold cross-validation scheme, where for each fold a new set of training and test trials were used.

RESULTS

Behavior

Overall performance on the first and second target tasks are shown as a function of T1-T2 lag in Figure 2a. Mean T1 AUC was 0.90 (SEM=0.04), and did not change as a function of temporal lag (F(1,12)=1.5, p>0.23). Mean T2 AUC was 0.72 (SEM=0.03), but unlike T1 performance, T2 performance showed an effect of lag, such that performance was much worse at the short T1-T2 lag than at the long lag (F(1,12)=21.4, p<0.001). Direct comparison of performance in the two tasks revealed that performance was impaired in the second task (F(1,12)=20.2, p<0.001) and that this impairment was worse at the short lags, as indicated by a significant task x lag interaction (F(1,12)=28.7, p<0.001). The impairment in T2 performance reflects the presence of the AB.

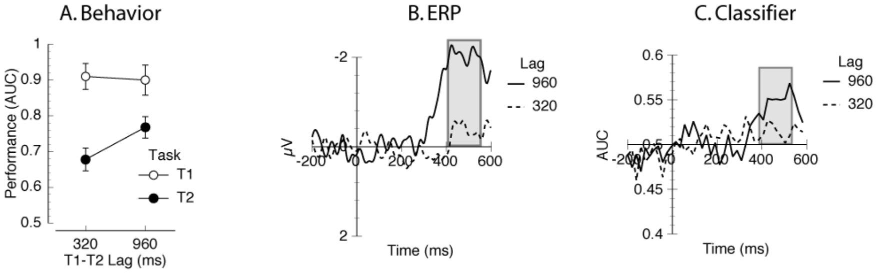


Figure 2. Panel A. Mean AUC on the T1 and T2 tasks plotted as a function of T1-T2 lag. Error bars represent ±1 standard error of the mean. Panel B. Grand average T2-evoked difference ERP wave forms plotted as a function of lag computed from midline electrodes Fz, Cz, and Pz. Shaded region indicates 400-500 ms, the period over which the mean amplitude statistics were computed. Panel C. Mean classifier accuracy based on T2-evoked activity recorded at the same electrodes plotted in B. Shaded regions indicates 400-500 ms time-window.

Electrophysiology

Two analyses of the ERP data were performed. The first focused on analyzing the data based on the type of stimulus presented to the observer. The second focused on analyzing the data based on whether the subject was correct or incorrect.

Analysis of stimulus type

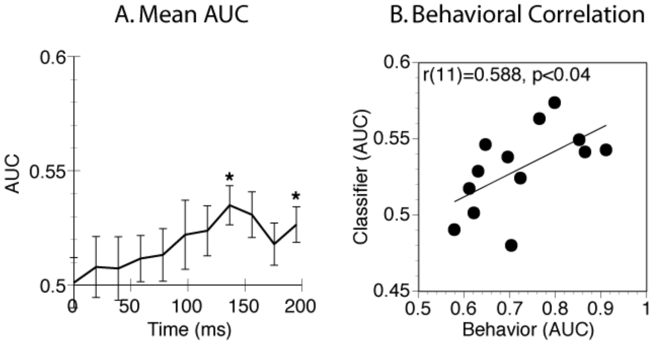
The results of the traditional ERP analysis are shown in Figure 2b. Shown are the average N400 difference waves from midline electrodes (Fz, Cz, Pz) on trials in which T2 was presented 320 ms after T1 and on trials in which T2 was presented 960 ms after T1. Analysis of the mean amplitude over the 400-500 ms time window (highlighted region) revealed a significant N400 at long lags (t(12)=4.12, p<0.002), but not at short lags (t(12)=1.19, p>0.23). These results suggest that access to semantic information was suppressed during the AB.

The results of the classification analysis are shown in Figure 2c. The overall pattern over time was qualitatively similar to that of the ERP data. Analysis of the mean classification accuracy over the 400-500 ms time window (highlighted region), paralleled the ERP results such that at the long temporal lag, classifier accuracy was significantly greater than chance (t(12)=2.60, p<0.03). At the short temporal lag, however, the classifier performance was not reliably different than chance (t(12)=1.18, p>0.26). These results indicate that information about the semantic relationship between T2 and the context word is represented in patterns of neural activity, even though overall performance is impaired relative to performance on the first task.

While these results are encouraging, it possible that the poor classifier performance during the AB reflects a problem with classifying stimulus information when performance is generally bad (i.e., during the AB). While this is plausible, it is probably unlikely because, even though performance at the 920 ms lag was better than at the 320 ms lag, it was still impaired relative to performance on the first task. Nevertheless, to rule out this possible shortcoming, we applied the same pattern classification analysis to a set of previously published data in which we observed a robust N400 during the AB (Giesbrecht, et al., 2007). Critically, mean classification during the same time window was 0.56 (SEM=0.019), which was significantly greater than chance (t(11)=3.23, p<0.02). Thus, when considered together, pattern classification algorithms can accurately discriminate the type of stimulus presented to the observer during the AB.

Analysis of performance failures

The second main analysis performed on the EEG data focused on classification of trials in which observers correctly discriminated T2 vs. trials in which they did not correctly discriminate T2. This analysis was aimed at whether this classification could be done based on activity evoked by the first target. The rationale was based on two premises: 1) the theoretical notion that the key determinant in triggering the AB is the extent to which subjects allocate attention to T1 and 2) detecting performance failures prior to their occurrence may provide a key advancement for the development of tools for online monitoring of attentional states. Because we focused on activity evoked by T1and because the 320 ms and 960 ms lag conditions were randomly intermixed, we averaged performance of the classifier across lags. The results of this analysis are shown in Figure 3a, which plots classifier accuracy in separate 20 ms time bins over the first 200 ms of T1-evoked activity. Classifier increased over time and was significantly above chance by 130 ms (mean=0.54, SEM=0.008, t(12)=4.10,p<0.002, corrected for multiple comparisons). To assess the relationship between T1-evoked classifier performance and individual differences in T2 performance, we correlated average T2 performance with the mean of the classifier performance at the time points that survived the Bonferroni corrected test versus change (indicated by the asterisks). The scatterplot shown in Figure 3b shows a significant positive correlation between behavioral performance and classifier performance (r(11)=0.588, p<0.04). These results demonstrate that activity evoked by T1 can be used to discriminate T2-accurate from T2-inaccurate trials prior to the presentation of T2.



A. Mean AUC

B. Behavioral Correlation

Figure 3. Panel A. Mean classifier AUC discriminating between T2-correct from T2-incorrect trials based on T1-evoked activity recorded at 64 electrodes. Error bars represent ±1 standard error of the mean. Asterisks represent time-points that are significantly different from chance, $p<0.05$, two-tailed, corrected for multiple comparisons. Panel B. Correlation between mean T2 performance and classifier performance at the time-points that survived the statistical threshold used in A.

DISCUSSION

The aim of the present work was to investigate the neural mechanisms of attentional failures by applying computational learning algorithms to measures of neural activity acquired while participants performed a difficult dual task. Our results demonstrated two key findings. The first key finding was that machine learning algorithms can be used to discriminate between two classes of stimuli presented to an observer during periods when behavioral performance is impaired (i.e., during the AB). While the successful discrimination between stimulus types is likely to be constrained by the inherent differences between the stimulus classes themselves, the finding of successful classification accuracy during the AB is consistent with the notion that information about the external world is represented in patterns of neural activity, even though conscious access to those representations may be impaired (e.g., Haynes & Rees, 2005; Kamitani & Tong, 2005; Philiastides & Sajda, 2006a, 2006b).

The second, and perhaps more important finding for the field of neuroergonomics, is that pattern classifiers can be used to discriminate between trials on which the observer correctly discriminated T2 vs. trials on which observers incorrectly discriminated T2. Critically, successful classification of these two types of trials was based on patterns of neural activity evoked by the T1. In other words, our analyses were able to discriminate between performance failures and successes based on neural responses that occurred more than 200 ms before the imperative stimulus and more than 1 second before the motor response to that stimulus. Moreover, classification accuracy was correlated with individual differences in performance. This finding suggests that pattern classification algorithms combined with continuous measurement of EEG response may be a viable tool for online monitoring of cognitive states in multiple performance contexts so that failures of attention can be detected when, and perhaps even prior to, their occurrence.

The present results converge with studies in the machine learning literature showing successful tracking of shifts of covert attention (Zhang, Maye, Gao, Hong, Engel, & Gao, 2010) and real-time interfacing between the brain and computer in untrained subjects (Blankertz, Losch, Krauledat, Dornhege, Curio, & Muller, 2008). The present finding that performance failures can be predicted before the imperative stimulus is presented suggest that it may be feasible to use similar online algorithms to interface with adaptive systems that monitor the user's attentional state and adjust display parameters to optimize performance. In such a system, the user could be alerted or the task altered based on the attentional load determined from online classification, a process that could prevent human performance errors before they occur.

ACKNOWLEDGEMENTS

This work was generously supported by the Institute for Collaborative Biotechnologies contract no. W911NF-09-0001 from the US Army.

REFERENCES

B. Blankertz, F. Losch, M. Krauledat, G. Dornhege, G. Curio & K.R. Müller. (2008) The Berlin Brain–Computer Interface: Accurate performance from first-session in BCI-naive subjects, *IEEE Transactions on Biomedical Engineering*, 55 (10), 2452–2462.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics, 7,* 179–188.

Giesbrecht, B., & Kingstone, A. (2004). Right hemisphere involvement in the attentional blink: Evidence from a split-brain patient. *Brain & Cognition, 55*(2), 303-306.

Giesbrecht, B., & Mangun, G. R. (2005). Identifying the neural systems of top-down attentional control: A meta-analytic approach. In L. Itti, G. Rees & J. Tsotsos (Eds.), *Neurobiology of Attention*. New York: Academic Press/Elsevier.

Giesbrecht, B., Sy, J. L., & Elliott, J. E. (2007). Electrophysiological evidence for both perceptual and post-perceptual selection during the attentional blink. *Journal of Cognitive Neuroscience, 19*, 2005-2018.

Giesbrecht, B., Woldorff, M. G., Song, A. W., & Mangun, G. R. (2003). Neural mechanisms of top-down control during spatial and feature attention. *Neuroimage, 19*, 496-512.

Haynes, J. D., & Rees, G. (2005). Predicting the orientiation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience, 8*, 686-691.

Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nature Neuroscience, 3*(3), 284-291.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contexts of the human brain. *Nature Neuroscience, 8*, 679-675.

Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature, 383*, 616-618.

Marois, R., Chun, M. M., & Gore, J. C. (2000). Neural correlates of the attentional blink. *Neuron, 28*(1), 299-308.

Marois, R., Yi, D.-J., & Chun, M. M. (2004). The neural fate of consciously perceived and missed events in the attentional blink. *Neuron, 41*, 465-472.

Philiastides , M. G., & Sajda, P. (2006a). Neural representation of task difficulty and decision making during perceptual categorization: a timing diagram. *Journal of Neuroscience, 26*, 8965-8975.

Philiastides , M. G., & Sajda, P. (2006b). Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cerebral Cortex, 16*, 509-518.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance, 18*, 849-860.

Rolke, B., Heil, M., Streb, J., & Henninghausen, E. (2001). Missed prime words within the attentional blink evoke an N400 semantic priming effect. *Psychophysiology, 38*, 165-174.

Slagter, H. A., Lutz, A., Greischar, L. L., Francis, A. D., Nieuwenhuis, S., Davis, J. M., et al. (2007). Mental training affects distribution of limited brain resources. *PLoS Biology, 5*, 1228-1235.

Vogel, E., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature, 428*, 748-751.

Zhang, D., Maye, A., Gao, X., Hong, B., Engel, A.K., Gao, S. (2010). An independent brain-computer interface using covert non-spatial visual selective attention. *Journal of Neural Engineering, 7*, 1-11.