What do Users Feel? Towards Affective EEG Correlates of Cybersecurity Notifications

Colin Conrad*, Jasmine Aziz, Natalie Smith and Aaron Newman[1]

[1] Dalhousie University, Halifax, Canada
*colin.conrad@dal.ca

Abstract. Security notifications attempt to change risky computer usage behaviour but often fail to achieve their desired effect. Though there are likely many causes for this phenomenon, information systems researchers have posited that emotional reactions to security notifications may play a role in its explanation. This work-in-progress paper descibes a study to create a baseline of electroencephalographic (EEG) and behavioral responses to security notification images by comparing them to known responses to the well-studied International Affective Picture System (IAPS). By creating such a baseline of affective responses to security notification images, future work can explore the effect of passive emotional reactions to security notification designs which would generate insight into effective design practices.

Keywords: Security warnings · Affective processing · Electroencephalography (EEG) · Event-related potential (ERP) · Late positive potential (LPP)

1 Introduction

Security notifications play an important role in the safe operation of computing environments, by informing users of threats and persuading them to change their computer behavior. Security notifications are particularly interesting to information systems researchers because they can fail to evoke desired behavioral change, for reasons that users may not be explicitly aware of. Though there has been considerable recent progress in the design of effective security warnings, future improvements to security information systems may be made by identifying contextual factors that influence secure behavior [1]. Affective considerations, such as degrees of trust, safety, or fear may play a role in unconsciously mediating the relationship between security notifications and behavior [1, 2].

Emotions can be conceptualized as either positive valence (e.g. joy or happiness) or negative valence (e.g. fear, anger or disgust), as well as by their degree of arousal. The late positive potential (LPP), an event-related potential (ERP), has been associated with both high and low valence emotions that also elicit high degrees of arousal [3, 4]. As such, the LPP can potentially be used as a marker of high-arousal emotional responses. In this work-in-progress paper, we describe an experiment to measure an association between the LPP and images of security notifications. Using the International Affective Picture System (IAPS) [5], we will compare LPP responses elicited by standardized negative, positive, and neutral valence stimuli, with those elicited by pictures of computer security notifications and security-unrelated computer images. Motivated by past studies which found that security warnings often fail to produce the desired reactions in users [6], we explore affective correlates of security warning pictures. The outcome of this study will be a baseline of LPP and questionnaire responses which can be used to passively and objectively investigate valence of novel security notification designs.

2 Background and Theoretical Framework

2.1 Security Notifications, Emotions, and Neurophysiology

Computer users are known to resist persuasion by protective messages—such as those given by security notifications—for reasons which they may not be explicitly aware [2]. Until very recently, information systems (IS) research on the subject has focused on either cognitive factors such as habituation or cognitive processing, or on negative emotional factors such as stress or fear [2]. Concerning fear, for instance, protection motivation theory has been identified as useful for explaining desktop security behavior [7-9]. As conceptualized by Rogers [7], this theory holds that there are at least four components that could determine a users' response to a threat: their perception of threat susceptibility, their perception of threat severity, their perception of response efficacy, and their perception of their personal ability to effectively respond [8]. Emotional reactions elicited by security notifications could thus influence perceptions of vulnerability and responses to threats, which ultimately influence behavior.

Though IS researchers have investigated affective factors in the processing of security notifications, much of the past research has been conducted using self-report measures [10, 11]. Such instruments are likely useful for measuring motivation, though they might not effectively measure implicit emotional responses to security notifications. Recognizing limitations to these studies, IS researchers have begun to employ neuroscientific and physiological techniques to investigate security phenomena [2, 9, 12, 13]. Such an approach promises to yield insights into unconscious affective factors which influence motivation for security behavior.

In a 2014 paper, Vance et al. [6] used a combination of EEG and questionnaire measures to predict disregard to security notifications. They found that an attention-related P300 ERP response to a gambling and risk task was a strong predictor of a participant's propensity to disregard security responses, when compared to questionnaire measures. Drawing from this study, we can expand on their findings by exploring a similar ERP measure (i.e., the LPP) to investigate affective factors that may influence this propensity. While Vance et al. [6] investigated attention-related ERP responses to gambling tasks to predict risky behavior, we instead investigate emotion-related ERP responses to security notifications themselves. By doing this, we may identify ERP measures which either better predict threat susceptibility than questionnaires or can later be applied as a passive, real-time measure in an ecologically valid setting.

2.2 EEG and the Late Positive Potential (LPP)

There is an extant literature on measuring emotional responses using ERP. Much of the literature has focused on two ERP components: an early posterior negative component at the 200 ms latency range, and a late positive component (LPP) which often starts at 300 ms and extends to 2000 ms [14, 15]. While the former is thought to affect processes related to cognition of emotions, the LPP may actually consist of an enlarged P300 component which extends well-beyond the normal latency of the P300 response, reflecting an effect of extended task-relevance to an emotion-inducing stimulus [14, 15]. Studies of the LPP have demonstrated an effect that is modulated by the strength of emotions evoked by pictures [16, 17].

Before investigating the impact of affective ERP correlates on security behavior, it is desirable to first have a baseline of responses to stimuli that have been standardized with respect to their emotional valence and arousal levels, for comparison to various security stimuli. The IAPS [5] is a well-studied repository of images which have been indexed based on normative ratings to emotion (valence, arousal, dominance) for study of attention and emotion. Though the IAPS is often used to investigate physiological processes elicited by emotions such as brain oscillations [18], the IAPS has also been used to investigate the emotional effect of art [19], and to validate the measure of emotions in a virtual reality environment [20]. In addition, the IAPS normative ratings are based on responses to the Self-Assessment Manikin (SAM), which is a well-studied affective rating questionnaire system. By combining both physiological and psychological approaches, we may discover gaps between security threat perceptions and unconscious physiological responses. Though we hypothesize that the selected security notifications will exhibit patterns characteristic of neutral photos, the guiding purpose of the study described in this paper is to identify a baseline of affective reactions to security stimuli. Our research question can thus be articulated as follows:

> RQ: How do the LPP and self-report measures of valence and arousal differ between security notifications, computer task images, and IAPS stimuli?

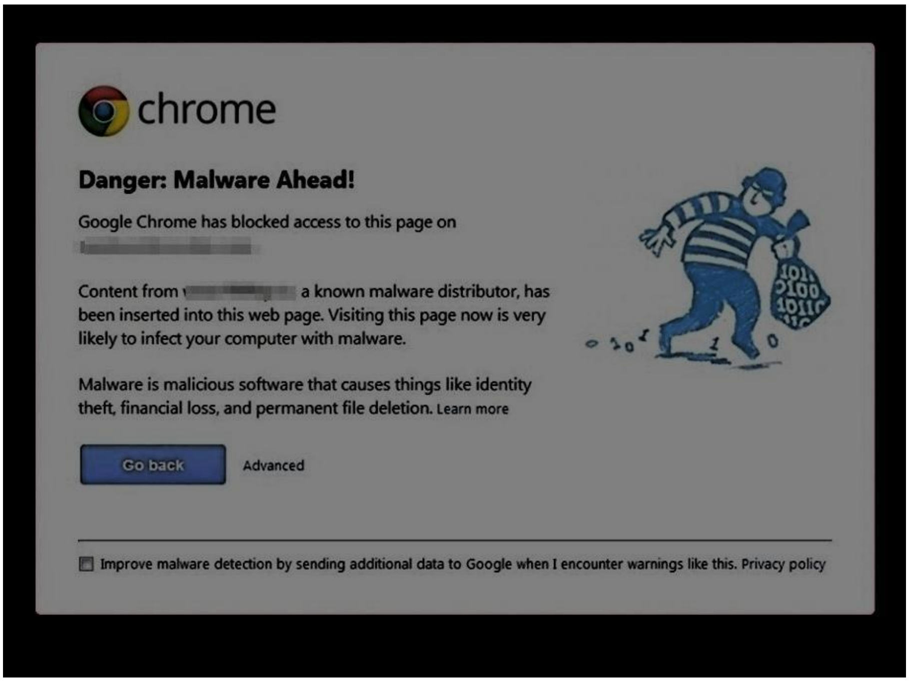
3 Methods

3.1 Participants

For the initial study, we will recruit 30 undergraduate students from our university who will be compensated with either $20 cash or course credit. Participants will be excluded if they reported having neurological conditions that could affect EEG (e.g. epilepsy or a recent concussion), uncorrected vision problems, or physical impairments that would prevent them from using a computer keyboard or mouse. In this work-in-progress paper, we report preliminary results from 3 participants.

3.2 Stimuli

Experimental stimuli will include 3 categories of photos from the IAPS database (negative, positive and neutral), as well as two categories of online computer stimuli (security notification images and neutral computer-related images) delivered following a within-subject design, presenting 32 instances of each of the five picture conditions. Security notification images will consist of computer-based images used by antivirus, web browser and firewall systems (e.g. Chrome, McAfee, Norton) while the computer-related images will consist of non-threatening images typical of a computing environment (e.g. a screenshot of a search engine or Wikipedia). All images will be corrected for luminance to control for the effect of luminance on EEG signals [21]. Stimuli will be delivered using PsychoPy [22, 23], which will also be used to mark the onset of each pictures in the EEG recordings via transistor-transistor logic (TTL) codes. A collection of both modern and antiquated security notifications were selected to give a greater range of baseline data.



Danger: Malware Ahead!

Google Chrome has blocked access to this page on [redacted]

Content from [redacted] a known malware distributor, has been inserted into this web page. Visiting this page now is very likely to infect your computer with malware.

Malware is malicious software that causes things like identity theft, financial loss, and permanent file deletion. Learn more

□ Improve malware detection by sending additional data to Google when I encounter warnings like this. Privacy policy

Fig. 1. A sample security warning stimulus. The effect of luminance on emotional processing will be controlled by normalizing all stimuli photos to the luminance baseline provided by the IAPS.

3.3 Questionnaires and Self-Assessment

At the outset of the experiment, participants will be asked about their age, gender, perceived skill at using computer systems, years of education and native language. To assess participants' perceived reaction to the photo stimuli, we will use a simplified version of the SAM [5, 24]. The manikin will be presented 2-3 seconds following the appearance of the stimulus photos and will consist of two 5-point scales which measure degrees of valence and arousal respectively [5]. Following the experiment, participants will be asked questions about their attitudes towards risk and perception of the impact of computer malware [6].

3.4 Procedure

Participants will undergo a consent protocol, will be fitted with an EEG cap, and then brought to a controlled environment. After participating in an initial demographic questionnaire, participants will be presented with a randomized series of images consisting of IAPS photos, security notifications or security-unrelated computer phenomena. Participants will complete the SAM measures for valence and arousal following each picture. Following the study participants will complete the aforementioned post-questionnaire. Each session is expected to take 90 minutes total. Following the session participants will be debriefed and will receive compensation.

3.5 Data Acquisition

Participants will be fitted with horizontal and vertical electrooculograms (EoG) and 32 scalp electrodes (ActiCap, BrainProducts GmbH, Munich, Germany) positioned at standard locations according to the international 10-10 system and referenced to the midline frontal location (FCz). Electrode impedances will be kept below 20 kOhm at all channel locations throughout the experiment. EEG data will be recorded using a Refa8 amplifier (ANT, Enschende, The Netherlands) at a sampling rate of 512 Hz, bandpass filtered between 0.01 and 170 Hz, and saved using ANT ASAlab.

3.6 Data Processing and Analysis

Data processing and statistical analysis will be conducted in Python using the MNE Python library [23, 25]. Data will be filtered using a 0.1-40 Hz bandpass filter and will be manually inspected for excessively noisy electrodes, which will be removed. The data will be segmented into 2200 ms epochs spanning from 200 ms before the stimulus onset, through the 2000 ms duration of the photo stimuli. Epochs will be manually inspected and those with excessive noise will be removed. Independent component analysis [26] will be used to remove systematic artifacts from the data, including those created by eye blinks, eye movements, and muscle contractions.

Each participant will yield a maximum of 32 epochs for each condition, and the dependent EEG measure will be mean amplitude on each trial between 300 ms and 2000 ms following stimulus outset, which corresponds to the expected window of the

LPP component. Statistical analyses will be performed on a region of interest centered around electrode Pz using linear mixed effects modelling [27-30]. Picture condition will be treated as fixed effects while participants, electrode-by-subject, and conditions-by-subject will be treated as random effects. The online security warnings condition will be selected as the fixed effect and the IAPS neutral condition will be specified as the intercept variable. Participants, participants-by-condition, and participants-by-electrode will be specified as random effects. Average SAM responses for each condition will be compared using ANOVA.

4 Preliminary Results

4.1 The LPP Waveform

Preliminary results suggest that there is variance in both EEG and behavioral measures between the conditions. To date, a total of 435 epochs have been analyzed from 3 participants, though collection was suspended due to the COVID-19 pandemic. Results from linear mixed effects analysis found a significant difference between the neutral IAPS and the positive inflection, presumably created by the LPP, elicited by pictures of the security warning condition ($\beta = 2.327$; $t = 2.38$; $p = 0.017$), as well as the positive IAPS condition ($\beta = 2.052$; $t = 2.12$; $p = 0.034$), and the negative IAPS condition ($\beta = 3.199$; $t = 3.28$; $p = 0.001$). Fig. 2 visualizes the grand average waveform for three of the five conditions.

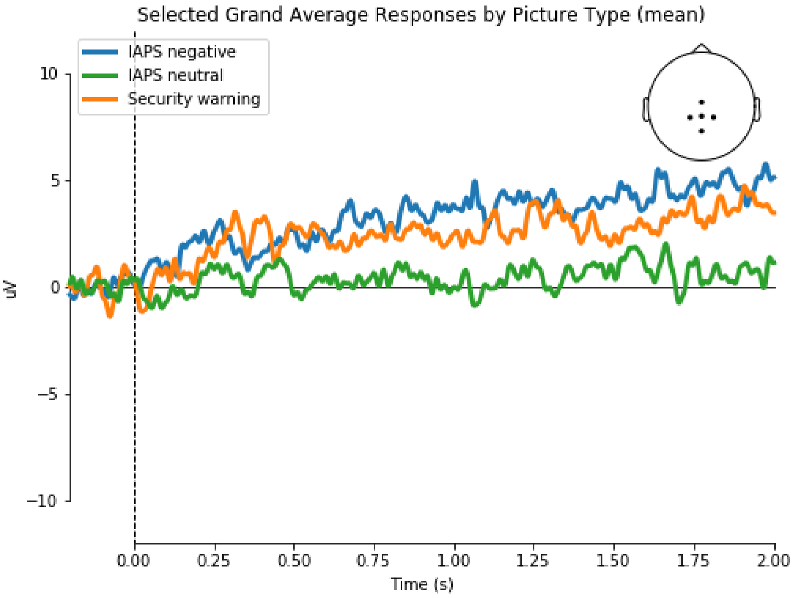


Fig. 2. Selected comparisons of LPP grand average waveforms for the region of interest. Three conditions were selected in order to enhance readability. Variances in response suggest the potential for an effect.

4.2 Valence and Arousal Reports

Preliminary results from ANOVA analysis of the SAM valence responses indicate a statistically significant effect of picture type on valence responses (F = 36.50; p < 0.001), though not arousal responses (F = 0.94; p = 0.48). Posthoc analysis using Tukey's test revealed significant differences between security warnings and neutral stimuli (p = 0.035), as well as between security warnings and positive stimuli (p = 0.001). Responses from computer-related pictures were found to be significantly different positive pictures (p = 0.011) but no other conditions. Fig. 3 summarizes the mean SAM valence rating responses and Fig. 4 summarizes the SAM arousal ratings.

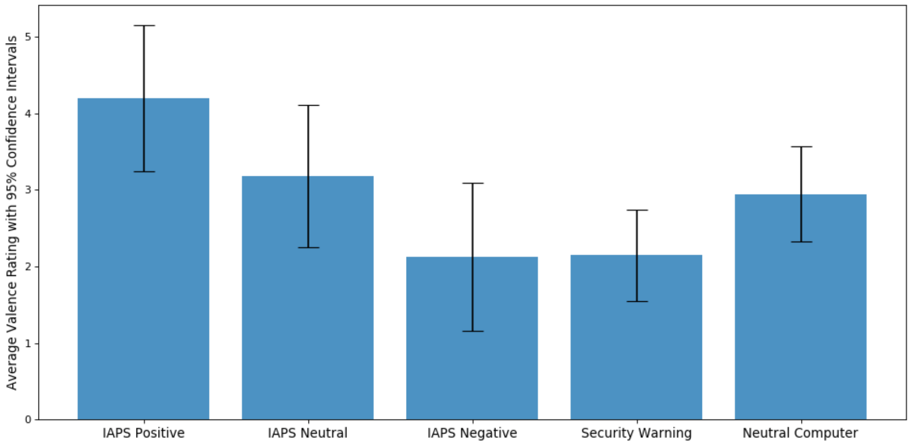


Fig. 3. Comparisons of valence responses from the SAM for each condition. Results from Tukey's test reveal significant differences in valence between IAPS positive stimuli and security warnings, as well as IAPS neutral stimuli and security warnings.

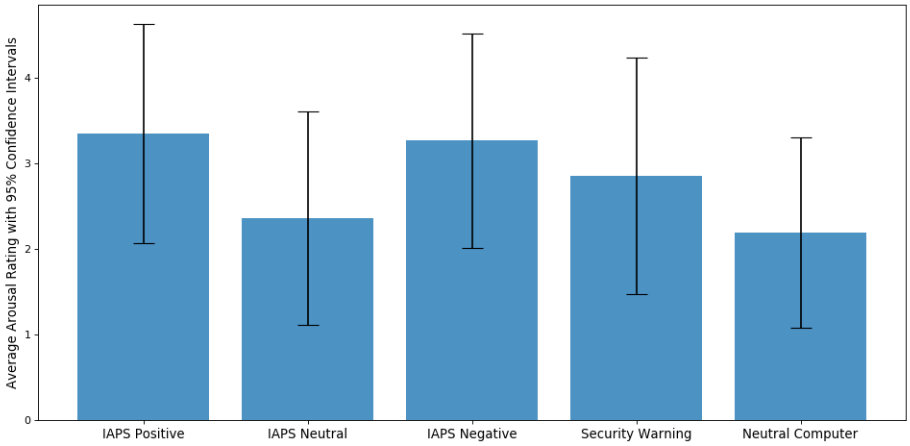


Fig. 4. Comparisons of arousal responses from the SAM for each condition. No significant results were found at this preliminary stage.

5 Discussion and Next Steps

Preliminary results revealed two interesting findings. The first is that there are differences in LPP between security warning stimuli and the IAPS neutral condition. Though the study does not yet have sufficient statistical power to draw conclusions, this suggests that that we may later gather evidence for this trend and may gain insights into differences between the conditions. The second finding is that the security warning condition was significantly different in reported valence from the positive and neutral IAPS conditions, but not the negative. If this effect holds true with greater statistical power, then we would have evidence to believe that the selected security warnings investigated are interpreted similarly to negative valence stimuli.

We anticipate two future research directions following the completion of this study. The first is to conduct a follow-up study of reactions to different varieties of security notifications (e.g. positive and encouraging notifications vs. fear-evoking notifications) or in different contexts (e.g. contexts of imminent negative consequences vs contexts of possible or future threats). In such a study, we could identify varieties of notifications to further investigate behavioral change outcomes, as well as the moderating effects of cognitive factors such as complexity. The second direction is to investigate the outcomes of responses to stimuli in ecologically valid settings. Similar to the study conducted by Vance et al. [6], future work could use deception to simulate an actual security risk and investigate the differences in reactions to positive and negative valence notifications and their effects on behavioral outcomes—although doing this effectively presents logistical and ethical challenges. Alternatively, such future challenges could be overcome by conducting this study in an office setting in co-operation with industrial partners. The present study nonetheless presents the first steps in extending the work done on security notifications in the information systems field towards a deeper investigation of affective brain processes.

References

1. Reeder, R. W., Porter Felt, A., Consolvo, S., Malkin, N., Thompson, C., & Egelman, S.: An experience sampling study of user reactions to browser warnings in the field. CHI ’18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 512, pp.1-13, doi: 10.1145/3173574.3174086 (2018)
2. Brinton Anderson, B., Vance, A., Kirwan, C. B., Eargle, D., & Jenkins, J. L.: How users perceive and respond to security messages: a NeuroIS research agenda and empirical study. European Journal of Information Systems, vol. 25, iss. 4, pp. 364-390 (2016)
3. Colombetti, G.: Appraising valence. Journal of Consciousness Studies, vol. 12, iss. 8-9, pp. 103-126 (2005)
4. Bublatzky, F., & Schupp, H. T. (2012). Pictures cueing threat: brain dynamics in viewing explicitly instructed danger cues. Social Cognitive and Affective Neuroscience, vol. 7, iss. 6, pp. 611–622 (2012)
5. Lang, P.J., Bradley, M.M., & Cuthbert, B.N.: International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8. University of Florida, Gainesville, FL (2008)

6. Vance, A., Anderson, B., Kirwan, C. B., & Eargle, D.: Using measures of risk perception to predict information security behavior: Insights from electroencephalography (EEG). Journal of the Association for Information Systems, vol. 15, iss. 10, no. 2 (2014).
7. Rogers, R. W.: Attitude change and information integration in fear appeals. Psychological Reports, vol. 56, iss. 1, pp. 179-182 (1983)
8. Hanus, B., & Wu, Y. A.: Impact of users' security awareness on desktop security behavior: A protection motivation theory perspective. Information Systems Management, vol. 33, iss. 1, pp. 2-16 (2016).
9. Warkentin, M., Walden, E., Johnston, A. C., Straub, D. W.: Neural correlates of protection motivation for secure IT behaviors: An fMRI examination. Journal of the Association for Information Systems, vol. 17, iss. 3, pp. 1940215 (2016).
10. Johnston, A. C., & Warkentin, M. (2010). Fear appeals and information security behaviours: An empirical study. Management Information Systems Quarterly 34(3), 549-556.
11. Guo, K. H., Yuan, Y., Archer, N. P., & Connelly, C. E.: Understanding nonmalicious security violations in the workplace: A composite behavior model. Journal of Management Information Systems, vol. 28 iss. 2, pp. 203–236 (2011)
12. Vance, A., Jenkins, J. L., Anderson, B. B., Bjornn, D. K., & Kirwan, C. B.: Tuning out security warnings: a longitudinal examination of habituation through fMRI, eye tracking, and field experiments. Management Information Systems Quarterly, vol. 42, iss. 2, pp. 355–380 (2018)
13. Kirwan, B., Anderson, B., Eargle, D., Jenkins, J., & Vance, A.: Using fMRI to Measure Stimulus Generalization of Software Notification to Security Warnings. In: Davis, F. D., Riedl, R., vom Brocke, J., Léger, P. M., Randolph A. B., Fischer, T. (eds.) Information Systems and Neuroscience. Lecture notes in Information Systems and Organisation, pp. 93-99. Springer International Publishing (2020)
14. Luck, S.: An Introduction to the Event-Related Potential Technique, 2nd Edition. (2014)
15. Hajcak, G., MacNamara, A., & Olvet, D. M.: Event-related potentials, emotion, and emotion regulation: an integrative review. Developmental Neuropsychology, vol. 35, iss. 2, pp. 129–155 (2010)
16. Hajcak, G., & Olvet, D. M.: The persistence of attention to emotion: brain potentials during and after picture presentation. Emotion, vol. 8, iss. 2, pp. 250–255 (2008)
17. Brown, S. B., van Steenbergen, H., Band, G. P., de Rover, M., & Nieuwenhuis, S.: Functional significance of the emotion-related late positive potential. Frontiers in human neuroscience, vol. 6, no. 33 (2012)
18. Güntekin, B., & Başar, E.: A review of brain oscillations in perception of faces and emotional pictures. Neuropsychologia, vol. 58, pp. 33–51 (2014)
19. Gerger, G., Leder, H., & Kremer, A.: Context effects on emotional and aesthetic evaluations of artworks and IAPS pictures. Acta Psychologica, vol. 151, pp. 174–183 (2014)
20. Marín-Morales, J., Higuera-Trujillo, J. L., Greco, A., Guixeres, J., Llinares, C., Scilingo, E. P., Alcañiz, M. & Valenza, G.: Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors. Scientific Reports, vol 8, iss. 1, pp. 1–15 (2018)
21. Bradley, M. M., Hamby, S., Löw, A., & Lang, P. J.: Brain potentials in perception: picture complexity and emotional arousal. Psychophysiology, vol. 44, iss. 3, pp. 364–373 (2007).
22. Peirce, J. W.: Generating stimuli for neuroscience using PsychoPy. Frontiers in Neuroinformatics vol. 2, no. 10 (2009).
23. Conrad, C., Agarwal, O., Woc, C. C., Chiles, T., Godfrey, D., Krueger, K., Marini, V., Sproul, A. & Newman, A. (2020). In: Davis, F.D., Riedl, R., Vom Brocke, J., Léger, P.-

M., Randolph, A., Fischer, T. (eds.) Information Systems and Neuroscience, pp. 287–293 (2020)

24. Leventon, J. S. & Bauer, P. J.: Emotion regulation during the encoding of emotional stimuli: Effects on subsequent memory. Journal of Experimental Child Psychology, vol. 142, pp. 312–333 (2016)
25. Gramfort, A., Luessi, M., Larson, e., Engemann, D. A., Strohmeier, D., Brodbeck, C., Parkkonen, L., Hämäläinen, M. S.: MNE software for processing MEG and EEG data. Neuroimage, vol. 86, pp. 446–460 (2014)
26. Delorme, A. & Makeig, S.: EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods, vol. 134, no. 1, pp. 9–21 (2004)
27. Davidson, D. J.: Functional mixed-effect models for electrophysiological responses. Neurophysiology, vol 41, iss. 1, pp. 71–79 (2009)
28. Newman, A. J., Tremblay, A., Nichols, E. S., Neville, H. J., and Ullman, M. T.: The influence of language proficiency on lexical semantic processing in native and late learners of English. Journal of Cognitive Neuroscience, vol. 24, iss. 5, pp. 1205–23. doi:10.1162/jocn a 00143 (2012)
29. Tremblay, A., and Newman, A.J.: Modelling non-linear relationships in ERP data using mixed-effects regression with R examples. Psychophysiology, vol 52, iss. 1, pp. 124–139 doi:10.1111/psyp.12299 (2014)
30. Conrad, C., & Newman, A.J.: Measuring the Impact of Mind Wandering in Real Time Using the P1-N1-P2 Auditory Evoked Potential. In Davis, F., Riedl, R., vom Brocke, J., Léger, P-M, & Randolph, A (Eds.). Information Systems and Neuroscience. Lecture notes in Information Systems and Organisation, pp. 37-45. Springer International Publishing doi: 10.1007/978-3-030-01087-4_5 (2018)