

Running head: EMOTIONAL RESPONSES TO PICTURE VERSUS WORD STIMULI

Event-Related Brain Potentials for Emotional Words versus Pictures: A Follow-up

Analysis of High-arousal Stimuli versus Low-arousal Stimuli

Lauren A. Smith

Roanoke College

Event-Related Brain Potentials for Emotional Words versus Pictures: A Follow-up Analysis of High-arousal Stimuli versus Low-arousal Stimuli

Over the course of biological evolution, the human race has developed and refined the most integral component of explicit communication: language. Using words to represent and symbolize specific referents in the environment around us has become a reliable and long-standing way to express ideas. Due to the importance and ubiquity of language in our everyday lives, it would be reasonable to assume that the human brain has structures devoted to the processing of words, much like other areas' roles in processing frequently encountered stimuli (i.e. the fusiform face area's primary role in facial processing; Kanwisher, McDermott, & Chun, 1997). But with the even older tradition of graphically representing the world through images (i.e. cave paintings), the question arises as to how these two modalities may differ. Is activation of brain areas involved in processing words and in processing pictures comparable? One way to assess this question is in using electroencephalography (EEG), a method that records the temporal aspect of brain activity in response to a time-locked stimulus (Weinberg & Hajcak, 2010). Although specified responses have been experimentally established in terms of neural activity using pictures through the use of EEG, fMRI, and PET technology (Weinberg & Hajcak, 2010; Kensinger & Schacter, 2006; Hajcak & Olvet, 2008; Mallan & Lipp, 2011; Schienle, Schäfer, & Naumann, 2008), a corresponding reaction to specific words representative of these images has been studied to a lesser extent. In light of the scarcity of evidence for trends associated with words *and* pictures of varying valence using EEG equipment, researchers Sara Dorrance, Tyler Stoneham, Kendra Boyd, Matt Bryant, Kayla Flippen, and I conducted a collaborative study exploring this issue. In addition to the group project, I continued on to explore further questions relating to the impact of arousal on an individual basis.

ERPs and Emotional Stimuli

In response to visual stimuli in the form of images, various event-related brain potentials (ERPs) have been shown to reliably indicate physiological responses indicative of emotional impact. ERPs are electrical potentials associated with specific sensory, perceptual, cognitive, or motor events that result from finding the time-locked average of EEG from many trials/presentations and then further reducing the background noise in order to isolate the signal (D. Nichols, lecture, September 10, 2013). More specifically, individual ERP components are “scalp-recorded neural activity that is generated in a given neuroanatomical module when a specific computational operation is performed” (D. Nichols, lecture, September 10, 2013). Two components in particular were identified based on their role ascertained from previous literature: P300 and late positive potential (LPP). According to Nechvatal and Lyons (2013), “The P300 peaks between 300 and 500ms following stimulus onset and has been linked to automatic attention processes, while the LPP typically peaks between 500 and 3000ms of stimulus onset and has been associated with controlled attention and emotion regulation” (p. 5).

Two measurements of emotionality were also considered in the design of the original study: valence and arousal. Valence has been referred to as how negative or positive a stimulus appears. Alternatively, arousal relates to how stimulating or calming an event is to the viewer (Kensinger & Schacter, 2006). Consistent with Schienle, et al (2008), P300 amplitudes correspond positively with emotionality such that higher amplitudes result from viewing pleasant and unpleasant stimuli. Neutral stimuli would therefore not be expected to elicit meaningful P300 amplitudes (Schienle, Schäfer, & Naumann, 2008). Based on previous research, LPP amplitudes indicate an increase in arousal as a response to pleasant and unpleasant stimuli when compared to those of neutral, low-arousal stimuli (Schienle, Schäfer, & Naumann, 2008).

Overall, research supports that valence is more closely associated with increased amplitudes in the P300 time ranges while arousal is more related to increased amplitudes in LPP time ranges. While the time ranges in which P300 and LPP amplitudes show significant magnitudes may vary across studies, we found that P300 components appeared between 290 ms and 390 ms after initial presentation of the stimulus while the LPP components appeared between 700 ms and 900 ms after initial presentation of the stimulus in our original study (**Figure 1**).

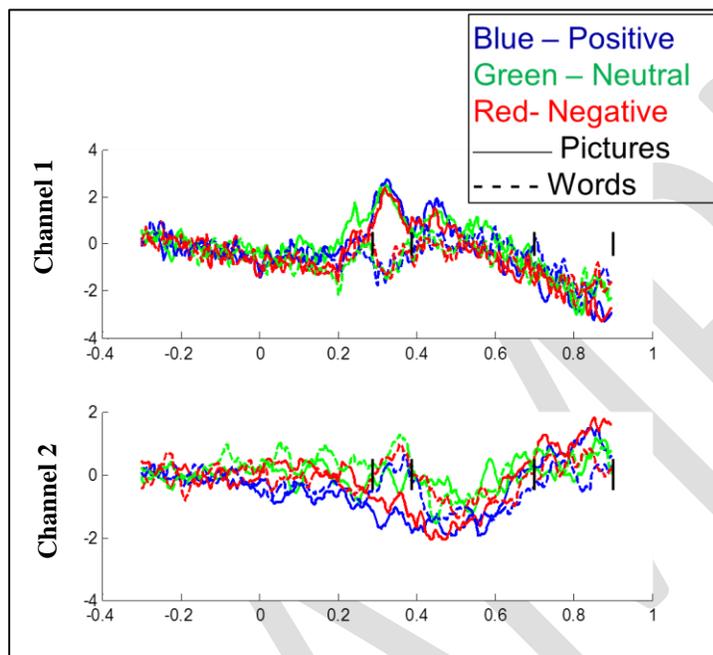


Figure 1 shows the different ERPs across time for each combination of stimuli and valence (e.g. positive word is a blue dashed line). The two sets of black vertical bars in each channel show the time ranges in which the ERPs were analyzed (dependent variables). The P300 component in both channels ranges from 0.29s to 0.39s and the LPP component in both channels ranges from 0.70s to 0.90s.

Citron (2012) reported that picture stimuli elicited higher brain activity, evidenced by ERPs, for emotion than word stimuli. These findings also confirmed that when looking at brain activity, both pictures and words show a distinction between valenced and neutral stimuli in the late phase between 500 and 800 ms after stimulus onset. However, in the early phase visual processing, between 200 and 300 ms after stimulus onset, the distinction was between arousing stimuli and neutral stimuli (Citron, 2012). It is important to note that these results had not yet been replicated. Further, Citron comments that effects of arousal have reportedly also been obtained in the late phase, without any distinction in terms of valence, when using EEG

equipment (2012). While the original study aimed primarily to discover a distinction between pictures and words at the P300 and LPP components, it inherently involved early and late phases as well (P300 and LPP respectively), with slightly different time windows.

After analyzing and applying concepts from B. F. Skinner, it is plausible that from an evolutionary standpoint, the natural human response to positive and negative stimuli could be necessary for survival. His theories suggest that our attention to negative stimuli arises from the fear of being harmed or killed, while our attention to positive stimuli arises from the need to seek companionship, have families, etc. This perspective implies that humans have a greater instinctual emotional response to those positive and negative stimuli necessary to live (Skinner, 1984). Skinner's concepts can be applied in understanding and explaining the results obtained from Citron (2012). Humans have always been able to see positive and negative visual stimuli for thousands of years, but have only recently begun reading.

Studies searching to identify physiological differences among words of varying valences have proven to yield mixed results. With intentional attention and processing to stimuli, emotional words elicit different ERPs than neutral words (Kissler, et al, 2007). Kissler, et al designed a study using "uninstructed reading," in which participants read text without directions to attend to certain words. These researchers found that emotionally charged words—both pleasant and unpleasant—did reveal increased amplitudes in target ERPs when participants were not given instructions (Kissler, et al, 2007). However, similar results have yet to be replicated in terms of viewing verbal stimuli with no instructions to direct attention to the semantic underpinnings of words presented (Kissler, et al, 2007).

Previous studies have shown that ERPs are enhanced for emotional pictures and words, but none have yet to consider related words and pictures in the same study. Thus, our previous

collaborative study examined the effects that type of stimulus and level of emotional valence had on brain activity. EEG equipment was used to record P300 and LPP activity of 21 undergraduate students. Two forms of stimuli (pictures and words) and three levels of emotionality (positive, neutral, and negative) were used. Analysis using a 3x2 Repeated Measures ANOVA showed that pictures generated higher magnitude P300 and LPP peaks when compared to words for all emotion levels. Also positive and negative stimuli generated higher magnitude P300 peaks and LPP peaks compared to neutral stimuli in a parietal electrode. Results from two channels were analyzed: Channel 1 consisted of electrodes on the fronto-occipital axis (one on the forehead, one on the back of the head located at theinion, and a ground on the forehead) while Channel 2 was comprised of a ground on the earlobe and another electrode above the centro-parietal area. According to Schupp, Junghöfer, Weike, and Hamm (2004), emotional pictures presented during EEG recording elicited higher LPP amplitudes over the centro-parietal area. Conversely, the lingual gyrus and fusiform gyrus of the occipital lobe tend to be primarily involved in the neural processing of words (Kuriki, Takeuchi, & Hirata, 1998). The results obtained from our study are very much in line with that of previous literature on similar topics.

Original Study

In accordance with the biological preparedness theory, which states that our arousal to certain stimuli is based upon experience and inheritance patterns (Citron, 2012), we expected to see greater brain activity for picture stimuli. Our results from the first study confirmed these expectations in that we saw heightened neural responsiveness to pictures compared to words in both the P300 and LPP components across both Channel 1 and 2 (Channel 1, P300:

$F(1,20)=72.466, p<0.001$; Channel 2, P300: $F(1,20)=23.321, p<0.001$; Channel 1, LPP:

$F(1,20)=22.201, p<0.001$; Channel 2, LPP: $F(1,20)=27.484, p<0.001$). **Figure 2** shows the results from P300 Channel 1.

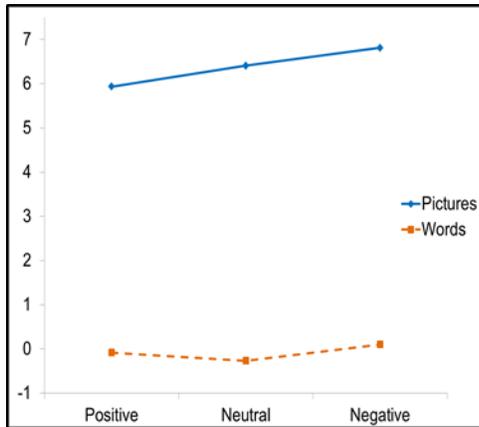


Figure 2. The graph to the left indicates the mean amplitudes expressed for emotional responses (positive, negative, and neutral) in response to pictures and words between the time ranges of .29s to .39s (P300) in Channel 1 (front-occipital axis). This graph shows a main effect of stimuli where pictures show greater brain activity than words. The lack of main effect for valence is indicative that emotional processing is not the primary role, which is in line with our predictions. The occipital lobe is involved more in visual processing, i.e. distinguishing between pictures and words on a basic level.

Despite the presumed hardwired predisposition we may have in response to pictures, words may have a unique encoding mechanism due to their symbolic nature; Kissler, Herbert, Peyk, and Junghöfer (2007) emphasize the learned nature of understanding words compared to pictures. While our original study did not take this idea into consideration, it became an important theory in light of the subsequent analysis.

We expected positive stimuli and negative stimuli to cause larger amplitudes in the P300 and LPP peaks in comparison to the neutral stimuli because positive and negative stimuli are expected to cause more emotionality and arousal in P300 and LPP components, respectively. What we found was a main effect of valence in which positive and negative stimuli elicited more brain activity than neutral stimuli, but this effect was only present in the LPP component (Channel 1, P300: $F(2,40)=1.391, p=0.261$; Channel 2, P300: $F(2,40)=3.125, p=0.055$); Channel 1, LPP: $F(2,40)=4.806, p<0.05$; Channel 2, LPP: $F(2,40)=8.386, p<0.05$). These results are shown in **Figure 3**.

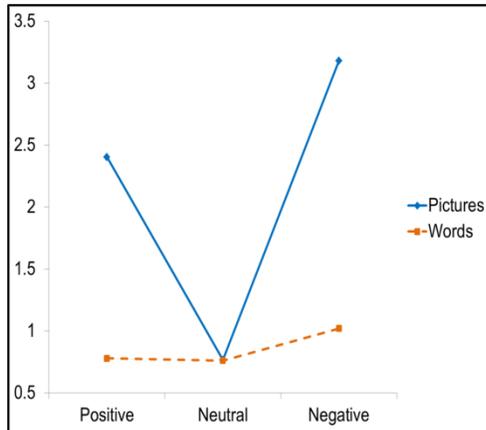


Figure 3. The graph to the left indicates the mean amplitudes expressed for emotional responses (positive, negative, and neutral) in response to pictures and words between the time ranges of .70s to .90s (LPP) in Channel 2 (centro-parietal lobe). This graph shows a main effect of stimuli where pictures show greater brain activity than words as well as a main effect of valence, where positive and negative pictures reveal significantly higher amplitudes than neutral pictures. The interaction effect is such that valence impacts physiological arousal of the stimulus only if the stimulus is in the form of a picture, but not for words. The nature of this interaction effect implies that emotional processing is happening, but on a more selective basis. While the valenced words are not drastically different from neutral words, the valenced pictures are very much different than the neutral pictures.

Individual Differences

A potential confound that may have emerged, however, as a result of grouping all participants' data is the inevitability of individuals' familiarity with certain stimuli due to previous experiences, which would directly impact their interpretation of and response to them. Certain stimuli that were agreed by the researchers to belong in the positive category might in fact induce fear in any number of participants, skewing the group averages. In effect, a wide range of emotions could be triggered based on previous exposure and experience with certain images due to individual differences. We cannot expect everyone to have the same emotional response to anything in reality, which emphasizes the need to standardize the subjective experience in some way.

Further, it should be noted that when looking back to **Figure 3**, it is clear that in the overall context of the stimuli there are distinct differences between positive and negative stimuli compared to neutral stimuli. These differences, however, reflect differences in the group average of all participants. By taking a closer look at each individual, it might be possible to identify a similar pattern on the individual basis.

Schupp, et al (2007) emphasized the role of P300 and LPP amplitudes contributing to selective attention to emotional stimuli. Their findings brought up a point of interest which was

pursued following the original study: increased P300 and LPP amplitudes corresponding to high-arousal stimuli (their categories included erotica and mutilation categories) when compared to the low-arousing controls (neutral category) (Schupp, et al, 2007). Our original results confirm the main effect of valence indicative of the presence of arousal, but do not reveal anything about the direction of this arousal. While this main effect is not significant in the P300 time range, it is significant in the LPP component across both Channel 1 and 2. These results reveal that both positive and negative stimuli were more stimulating than neutral stimuli, but this begs the question of how positive and negative stimuli relate to each other. Could it be that though they appear to reflect comparable levels of arousal (**Figure 3**), what we're really seeing is skewed due to the averaging of all pictures that comprise that category (i.e. all positive pictures)? No two positive pictures will elicit the same physiological response; some pictures may elicit a powerful response in the form of a large amplitude, while others may be less arousing to a specific individual. Further, amplitudes corresponding to positive and negative stimuli may be significantly higher compared to those associated with neutral stimuli, but there's no way of knowing if they are different from each other or not unless they are separated into high- and low-arousal subgroups. It is also worth noting that, as Citron (2012) mentions, "some studies report significant correlations between valence and arousal to some extent, i.e., the more highly valenced (positive or negative) a stimulus, the higher its level of arousal" (p. 211).

Current Study

The current study intends to take the findings of Schupp, et al (2007) and Citron (2012) into consideration in its further analysis of the results obtained from the original collaborative study. The new analysis more closely approximated the categories established by Schupp, et al, (2007) but shied away from the extremes of erotica and mutilation. In place of using these more

specific and predetermined “high-arousing” categories, this analysis will look to individuals’ ratings to identify what the participants considered to be relatively high-arousing or low-arousing. This is an important distinction that should be made in order to account for individual differences. Since the stimuli were originally placed into categories according to the researchers—and there was mild disagreement even among researchers as to whether or not a stimulus adequately embodied the category it was placed in—we cannot expect that each participant would view the selected stimuli as fitting into the categories to which they were designated. Due to this restructuring, the “high-arousal” category was comprised of those stimuli that each participant considered to be the most positive and most negative. Similarly, the “low-arousal” category will be determined by participants’ ratings of least positive and least negative stimuli.

Through this continued analysis, we hope to reveal a relationship between high-arousal stimuli and low-arousal controls, based upon participants’ subjective ratings, to brain activity. After determining how individual participants rated each of the positive and negative stimuli, we identified, on an individual basis and on a group basis, those stimuli which were rated most positive and most negative compared to those that were rated least positive and least negative. The stimuli rated most positive and most negative comprise the “High-arousal Stimuli” category, while those rated least positive and least negative make up the “Low-arousal Stimuli” category for each subject. The specific stimuli in each of these categories will vary from participant to participant. Subsequently, ERPs for each stimulus in these categories for each participant were identified and averaged across blocks.

In light of existing literature responding to similar questions and the efforts made to ameliorate any confounds, I expected to see a distinction in the LPP component. Specifically I

expected those items falling into the High-arousal Stimuli category to elicit significantly higher amplitudes in the LPP component when compared to those in the Low-arousal Stimuli category. In terms of how stimuli were subjectively rated, I did not expect to see a significant difference between the numbers of stimuli rated high or low within the categories established. In order to get a better idea of how the number of valid ERPs across participants compares between the High-arousal Stimuli and Low-arousal Stimuli, analysis was run on the valid EEGs as well. I would expect there to be minimal statistical significance since a significant difference between valid trials for High-arousal versus Low-arousal Stimuli would skew the data due to a lack of reliable data from which to draw upon.

Material and Methods: Original Study

Participants

A total of 21 participants (7 males, 14 females) were recruited from the student body at Roanoke College in Salem, VA for the original study. The participants included students from ages 18-23 years. All participants were required to have normal or corrected-to-normal vision and could not participate if susceptible to seizures induced by flashing stimuli. The experiments were conducted in accordance with the guidelines of the Roanoke College Institutional Review Board, with subjects providing informed consent.

Equipment

Electroencephalography (EEG) signals were recorded using a PowerLab 26T device from AD Instruments. Five lead shielded electrodes transmitted voltage signals from the scalp of the participant. The ground, Fp1, and right earlobe electrodes were disposable and simply stuck to the skin with their adhesive backing. The remaining electrodes, Oz and Pz, were attached to participants' scalps using electrode paste. All electrodes connected to a bio amp specially

designed to record signals in the biologically relevant range and to minimize artifacts from other electrical devices in the room. The room was designated in consideration of reducing artifacts. The analog input from the electrodes was converted by the PowerLab 26T into a digital time series output that is sent to a computer for additional processing by LabChart 7 software (AD Instruments, Inc). The time that stimuli were presented was indicated by a signal sent from an external Cedrus StimTracker device to the same computer, which was also recorded by the LabChart 7 software. The LabChart 7 software was run on a Dell XPS 15z laptop computer and presented on an external 17'' Dell monitor to be viewed by the experimenters only. The stimuli were presented on the internal 15'' widescreen monitor of the laptop using SuperLab 4.5.

Stimuli

Stimuli consisted of a total of 120 trials, comprised of 60 pictures and 60 words, which were displayed in random order. Unanimous personal reactions of all experimenters to each picture determined which grouping each would fall under: positive, neutral, or negative. The pictures, selected from the internet, consisted of 20 pictures that elicited positive emotions, 20 pictures that elicited neutral emotions, and 20 pictures that elicited negative emotions. After determining which pictures would be used, we chose one or two words to capture each picture, resulting in 20 positive words, 20 neutral words, and 20 negative words that corresponded directly with the pictures selected.

Each picture was displayed in grayscale in 250x250 pixels, while the words appeared in white font in size 24 against a gray background. Within the stimuli pictures of houses and the word 'house' appeared 8 times each. When a house stimulus was shown, the participant's task was to blink. **Figure 4** provides a sample stimulus from each category, along with the participants' subjective ratings of them.

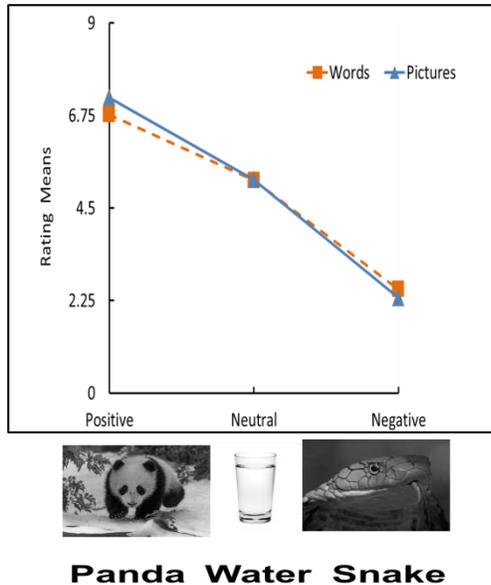


Figure 4 shows the mean subjective ratings of each stimulus. Subjects found words and pictures that were supposed to be positive as predominately positive, neutral as being neutral, and negative pictures and words were seen as negative. On the x-axis a picture and its corresponding word are shown representing each category. The panda comes from the positive subset, water from the neutral subset, and snake from the negative subset.

There were a total of 136 trials per block. Each trial was presented for 500 msec and there was an interval of 1000 to 1200 msec between trials, with the time between trials randomly chosen to avoid a constant interval in relation to the ongoing EEG oscillations. On average there was 1500 to 1700 msec from the start of one trial to the start of the next trial, making the total duration of a run approximately 4 minutes. There were 5 blocks of the 136 trials, making the total experiment roughly 20 minutes.

This portion of the study was followed by the subjective rating of all stimuli. Images and words were displayed on the screen for approximately the same amount of time as in the study. Participants were asked to rate each of the 60 pictures and 60 words, along with 2 of the task stimuli (1 picture of a house and the word “house”). Participants used the keyboard to select numbers from a range of 1 to 9, with 1 being the least pleasant, 5 being neutral, and 9 being the most pleasant.

Procedures

This study recorded amplitudes using five electrodes, measuring two channels. Channel 1 consisted of two electrodes: one at the anterior pole (Fp1) and one at the posterior pole (Oz) of

the fronto-occipital axis. Channel 2 was comprised of an electrode on the right earlobe and one above the centro-parietal area (Pz). The fifth electrode was the ground, placed on the forehead next to the Fp1 electrode.

Potential participants were presented with an informed consent sheet and pre-screened when they entered the experiment room. The pre-screening ensured that participants were not susceptible to seizures, had normal or corrected-to-normal vision, and did not have metal in their head. They were also warned about the possibility of a migraine headache as a result of viewing the stimuli.

After giving informed consent, participants abraded their skin where the electrodes were to be placed. Once participants cleaned the abraded areas with alcohol swabs, experimenters attached electrodes to the participants' scalps. A total of 5 lead shielded electrodes were used, an anterior-posterior comparison for Channel 1 (Fp1 vs. Oz), a comparison for Channel 2 using the posterior centro-parietal axis and earlobe as the two points (Pz vs. earlobe), and a ground on the forehead. All electrodes were held in place by elastic headbands and, additionally, the anterior, ground, and earlobe electrodes were kept in place by disposable electrodes. The locations for the anterior, posterior, and ground were chosen based on anatomical locations. The anterior and ground electrodes were placed symmetrically around the center of the forehead approximately two inches below the hairline. The posterior electrode was placed one inch above the inion (the bump on the back of the head). A headband was wrapped around the forehead and back of the head to hold in place the anterior and posterior electrodes. The earlobe and medial electrodes were placed in a practical fashion in relation to the other electrodes, though constrained by anatomical features. The medial electrode was placed dorsal and anterior on the scalp in relation

to the placement of Oz, beneath a headband that was wrapped below the chin and over the top of the head. The remaining electrode was placed on the right earlobe with a disposable fastener.

Prior to the start of data collection, the electrode connections were tested by viewing the output of the channels. The output of the channels was band pass filtered with an acceptable range from 0.5 to 50 Hz and a Mainz digital filter was applied. It was verified that the EEG signal for each channel remained within the range $\pm 60 \mu\text{V}$ when the participant held their head and eyes still and did not blink. During data collection, the signal was sampled at a rate of 400/s. A pillow was placed on a stack of books for a chin rest to aid the participant in keeping their head still. The size of the stack of books varied, ensuring that the participant was comfortable. Each participant completed 5 blocks, with each run lasting approximately 4 minutes. The participant had the chance to take a break in between each run; the runs were continued when the participant was ready to resume. After the completion of all runs, the electrodes were removed, cleaned, and the participant was debriefed.

Following completion of the experiment, the participant was given an electronic questionnaire in which they rated their general emotions toward the stimuli that they saw. This portion of the experiment was run on the laptop using a SuperLab program. The purpose of the experiment and the expected results were explained to the participant after the completion of the questionnaire. For most participants the total duration was approximately 40 minutes, including the set-up, body of the study, time for breaks, and questionnaire. The EEG data collected was then analyzed using MatLab and SPSS.

Materials and Methods: Current Study

Participants

For further analysis, participants whose EEG recordings yielded large amounts of background noise with indiscernible ERPs were thrown out. Following the identification of this data, remaining participants were narrowed down further in some cases. Data for participants that display a lack of variability in their subjective ratings were also eliminated. Due to this lack of variability in ratings, responses from these participants will more than likely be impossible to categorize into High-arousal Stimuli and Low-arousal Stimuli. As a result, the current study consisted of data from 11 participants (4 males, 7 females) between the ages of 18-23 years from the Roanoke College student body.

Equipment

Analysis was conducted using SPSS software and Matlab programming on the same laptops used to conduct the experiment.

Stimuli

All of the stimuli used in this analysis were derived from the original study, but all of the original stimuli were not necessarily analyzed. Regardless of how many were employed in the analysis, all of the stimuli came from the images and words from the positive and negative subgroups only; all neutral stimuli and task stimuli were excluded from additional analysis. Specific ERPs analyzed varied across participants, depending on which stimuli they rated as most positive, most negative, least positive, and least negative.

In determining which stimuli would be analyzed for each participant, two groups were formed: Group grouping and Individual grouping. In the Group grouping condition, anything that was rated as either a 7, 8, or 9 within positive stimuli and anything rated as a 1, 2, or 3 within negative stimuli were combined to create the High-arousal Stimuli subgroup. Those stimuli rated as either a 4, 5, or 6 within positive and negative stimuli comprised the Low-arousal

Stimuli subgroup within the Group groupings condition. The size of each subgroup then, was dependent upon the variability amongst individuals' ratings. If, for example, an individual rated every negative stimulus as a 1, 2, or 3 and every positive stimulus as a 7, 8, or 9, then that participant would have no stimuli in the Low-arousal Stimuli subgroup. This is a uniform approach ensuring that every stimulus in the Overall High condition had the same range of ratings (of either 1, 2, 3, 7, 8, or 9) and that every stimulus in the Overall Low condition also had the same range of ratings (of either 4, 5, or 6). The tradeoff with using this classification system is that each individual does not necessarily have the same amount of data as the others.

In the Individual groupings condition, the five most positive ratings in the positive stimuli group comprised part of the High-emotional Stimuli, along with the five most negative ratings in the negative stimuli group. Each subset could therefore include more than five stimuli and their corresponding ERPs. If, for example, an individual's most negative ratings include 2 ones, 2 threes, and 4 fours, the number of ERPs analyzed for that participant in the subcategory would be eight, before determining the number of positive ratings which would combine to form the High-arousal Stimuli subgroup. Although the first five of those would only include the 2 ones, 2 threes, and 1 of the fours, it would be impractical to assume that only one of the stimuli given the rating of four was any less negative than the remaining stimuli with the same rating. The same methodology determined which stimuli fell into the category of least positive and least negative, which comprised the Low-arousal Stimuli group. This individualistic approach ensures that each participants' subjective opinion was considered in relation to the rest of their ratings (i.e. some participants didn't use 1 to rate any stimuli, but a 3 would still be the most negative in relation to their 8). The drawback, however, is that some data had to be thrown out (i.e. if the participant

rated five of the twenty items as a 1, five of them as a 5, and the remaining ten as anything in between, those middle ten items were not used in compiling that participants' data).

Procedures

Further analysis was conducted using text documents recorded by the SuperLab program during the experiment. Using these text documents, we were able to determine how participants rated stimuli in the questionnaire as well as which order the stimuli were presented in for each of the five blocks for every participant. The setup of the SuperLab program enabled us to match comment markers to individual stimuli. Additionally, MatLab was used to average participants' ERPs across the five blocks for each category. SPSS software was integral in analyzing data and statistics in order to move forward and make conclusions. In addition to analyzing the P300 and LPP components, SPSS was used to discern whether or not significant differences between high and low ratings existed and further to determine if the percentage of valid trials varied significantly from expectations.

Analysis

Ratings. The difference between rating frequency was analyzed in order to ensure that this potential discrepancy was not contributing to the lack of significant results. If the analysis revealed a significant difference between the number of stimuli rated in the High-arousal group and the Low-arousal group, then this could be accountable. In order to run this analysis, we first coded ratings with either a 1 (High-arousal), a 2 (Low-arousal), or a 0 (neither category) according to the guidelines established for both Group grouping and Individual grouping classifications discussed earlier. See **Table 1** for an example of what the outcome of this process looked like for one participant in the Positive Pictures condition.

Category	Stimulus	Rating	Group grouping	Individual grouping
Stim1,	9	6	1	1
Stim1,	4	6	1	1
Stim1,	16	6	1	1
Stim1,	17	6	1	1
Stim1,	19	6	1	1
Stim1,	7	6	1	1
Stim1,	18	7	2	0
Stim1,	12	7	2	0
Stim1,	13	7	2	0
Stim1,	20	7	2	0
Stim1,	14	7	2	0
Stim1,	10	7	2	0
Stim1,	8	7	2	0
Stim1,	3	7	2	0
Stim1,	5	8	2	2
Stim1,	2	8	2	2
Stim1,	6	8	2	2
Stim1,	1	8	2	2
Stim1,	11	8	2	2
Stim1,	15	9	2	2
Total High			6	6
Total Low			14	6

Table 1 provides an example of how stimuli were coded according to the Group groupings and Individual groupings classification systems. This sample comes from a participant's ratings of the positive pictures (Stim1) for all 20 stimuli in that category. This particular participant relied on ratings 6, 7, 8, and 9. While under the Group groupings system there were more than twice as many ratings coded as Low compared to High, the number of each was equal in the Individual groupings system.

Following this step, the frequency of those stimuli coded with a 1 was totaled, as well as those stimuli coded with a 2, for each participant within each of eight conditions (valence: low, high; stimulus: picture, word; classification: group groupings, individual groupings). See **Table 2** for a summary of these results. SPSS software was then used to determine whether any of the differences within categories were significant.

	Group grouping				Individual grouping			
	Low_pic	Hi_pic	Low_word	Hi_word	Low_pic	Hi_pic	Low_word	Hi_word
Part. 1	19	21	24	16	16	24	16	16
Part. 2	7	32	13	26	9	10	10	17
Part. 3	13	26	23	17	13	13	16	17
Part. 4	10	30	16	24	16	17	19	15
Part. 5	19	21	24	16	14	18	18	15
Part. 6	20	19	5	34	24	15	37	2
Part. 7	7	33	14	26	15	17	14	15
Part. 8	9	31	8	32	14	18	21	19
Part. 9	6	33	14	26	11	24	16	16
Part. 10	5	33	10	28	14	23	14	23
Part. 11	9	29	16	23	14	17	16	18
Average	11.27	28.00	15.18	24.36	14.55	17.82	17.91	15.73
SE	1.69	1.62	1.93	1.83	1.13	1.34	2.10	1.54

Table 2 summarizes the number of stimuli falling into each category (Low- versus High-arousal) across all participants for each stimulus type (pictures and words) according to both classification systems (Group groupings and Individual groupings). On average, there was somewhat of a polarization within the Group grouping where high pictures and words had a greater number of stimuli compared to low pictures and words. This effect doesn't seem to hold up when looking at the Individual groupings classification. In order to determine whether or not the difference observed within Group groupings was significant enough to impact our results, we ran repeated measures ANOVA analysis.

Valid ERPs. Since the difference between ratings revealed no significant interference with the overall findings, we subsequently ran an analysis on the percentage of valid ERPs within individuals in order to discover if this might have impacted the results. This number was ascertained by looking first at the ratings; if after the coding of High- and Low-arousal across Group groupings and Individual groupings, an individual had 7 stimuli, for example, in the Low-arousal/pictures/Group groupings category, that number was multiplied by 5. We multiplied the number of stimuli in each category for each participant by 5 because there were 5 blocks, resulting in 5 presentations of that specific stimulus, over the course of the experiment. So in this case, 35 would have been the total number of possible valid ERPs in that category for that participant, assuming that the ERPs for every stimulus in that category across all trials were valid

and discernable. By looking at the MatLab output for each participant, we then determined how many ERPs out of that potential total were in fact valid and discernable. Out of the 35 potential ERPs, then, there may have been only 17 valid ERPs, resulting in 48.57% valid ERPs within that category for that participant. The total percentages for all participants were then analyzed using SPSS software.

Results

Ratings

As was predicted, a 2x2 (stimulus: pictures, words; valence: high, low) Repeated Measures ANOVA design yielded few significant results. Within the Group groupings category, there was a main effect of valence ($F(1,11)=20.093, p<0.05$) such that ratings for pictures used more of the scale; low pictures were rated lower within the Low-arousal Stimuli subgroup and rated higher within the High-arousal Stimuli compared to words. However, there was no main effect of stimuli present ($F(1,11)=3.750, p=0.082$), nor was there an interaction effect ($F(1,11)=3.447, p=0.093$) (**Figure 5**).

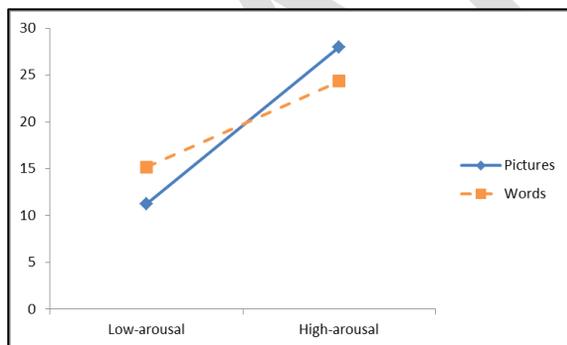


Figure 5 shows the frequency of stimuli falling into the categories of Low-arousal and High-arousal for pictures and words in the Group groupings classification. Overall, pictures received lower ratings in the Low-arousal subgroup and higher ratings in the High-arousal subgroup when compared to words.

Within the Individual groupings category, there was no significant main effect of stimulus ($F(1,11)=0.707, p=0.420$), no main effect of valence ($F(1,11)=0.050, p=0.827$), and no interaction effect ($F(1,11)=4.479, p=0.060$) present. The overall lack of significance suggests that stimuli were within the same range as others (**Figure 6**).

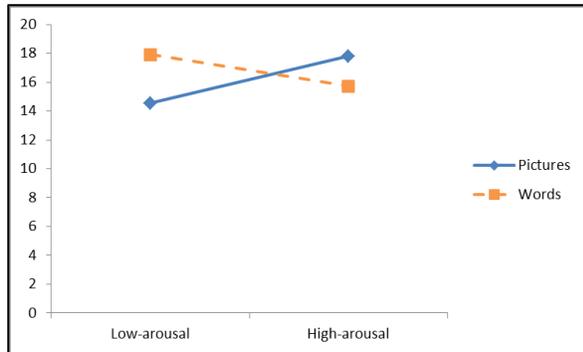


Figure 6 indicates the total frequency of stimuli in each category (Low-arousal and High-arousal) for both pictures and words in the Individual groupings classification. These results reveal no significant findings.

Valid ERPs

Repeated Measures ANOVA analysis confirmed the hypothesis that there was no significant difference between the percentages of valid ERPs across participants in the High-arousal and Low-arousal Stimuli subgroups. Within the Group groupings classification, data revealed no significant main effects of stimuli ($F(1,11)=3.564, p=0.088$) or valence ($F(1,11)=0.003, p=0.960$). There was also no interaction effect present ($F(1,11)=2.173, p=0.171$). Channel 2 revealed similar results (main effect of stimuli: $F(1,11)=1.753, p=0.215$; main effect of valence: $F(1,11)=0.775, p=0.399$; interaction effect: $F(1,11)=2.926, p=0.118$), indicating no significant results (**Table 3**).

	Channel 1				Channel 2			
	Low_pic	High_pic	Low_word	High_word	Low_pic	High_pic	Low_word	High_word
Part. 1	48.57	38.18	47.14	50.77	40.00	33.33	44.29	46.92
Part. 2	72.63	72.38	76.67	71.25	71.58	71.43	75.83	71.25
Part. 3	71.43	70.00	73.85	66.15	71.43	69.38	73.85	64.62
Part. 4	87.69	92.31	89.57	94.12	83.08	79.23	81.74	78.82
Part. 5	56.00	58.67	65.00	70.83	56.00	58.00	63.75	70.00
Part. 6	68.42	62.86	60.83	76.25	68.42	60.95	60.83	73.75
Part. 7	60.00	50.32	50.00	53.75	40.00	26.45	25.00	34.38
Part. 8	93.33	95.76	88.57	90.00	93.33	93.33	85.71	90.00
Part. 9	56.00	53.33	66.00	57.14	56.00	47.27	62.00	50.71
Part. 10	55.00	54.74	60.00	65.29	53.00	53.68	52.00	59.41
Part. 11	73.33	72.41	72.50	77.39	71.11	71.03	72.50	72.17
Average	67.49	65.54	68.19	70.27	64.00	60.37	63.41	64.73
SE	4.25	5.26	4.16	4.17	5.05	5.93	5.36	4.77

Table 3 shows the total percentage of valid trials for participants across Channels 1 and 2 for each category within the Group groupings classification.

Analysis of the Individual groupings classification similarly yielded no significant results for Channel 1 (main effect of stimuli: $F(1,11)=0.037, p=0.852$; main effect of valence: $F(1,11)=0.020, p=0.891$; interaction effect: $F(1,11)=1.474, p=0.253$) or for Channel 2 (main effect of stimuli: $F(1,11)=0.052, p=0.824$; main effect of valence: $F(1,11)=1.804, p=0.209$; interaction effect: $F(1,11)=0.173, p=0.686$) (**Table 4**).

	Channel 1				Channel 2			
	Low_pic	High_pic	Low_word	High_word	Low_pic	High_pic	Low_word	High_word
Part. 1	40.00	42.35	47.14	50.67	36.00	34.12	44.29	48.00
Part. 2	71.25	73.33	75.00	71.25	70.00	72.50	75.00	71.25
Part. 3	75.56	88.00	74.00	64.71	75.56	86.00	74.00	63.53
Part. 4	87.69	95.38	88.75	94.12	83.08	78.46	80.00	78.82
Part. 5	55.00	60.00	66.32	69.33	53.75	60.00	65.26	68.00
Part. 6	68.57	63.33	66.67	62.67	68.57	61.11	66.67	61.33
Part. 7	57.14	55.56	50.48	55.79	31.43	30.00	28.57	36.84
Part. 8	94.55	95.83	90.00	86.25	94.55	92.50	87.50	86.25
Part. 9	58.57	50.43	61.43	60.00	58.57	41.74	55.71	53.91
Part. 10	54.17	56.00	65.41	50.00	53.33	53.33	59.46	40.00
Part. 11	72.86	75.29	72.50	76.67	71.43	72.94	72.50	72.22
Average	66.85	68.68	68.88	67.40	63.30	62.06	64.45	61.83
SE	4.79	5.52	4.05	4.24	5.74	6.25	5.08	4.75

Table 4 shows the total percentage of valid trials for participants across Channels 1 and 2 for each category within the Individual groupings classification.

P300 Analysis

Upon analyzing the P300 amplitude with 2x2 (valence: High-arousal and Low-arousal; stimuli: pictures and words) Repeated Measures ANOVA, data revealed few significant results.

In the Group groupings classification in Channel 1, data revealed a main effect of stimulus

($F(1,11)=43.002, p<0.001$), in that pictures consistently elicited higher amplitudes than words.

There was, however, no main effect of valence ($F(1,11)=0.561, p=4.71$). An interaction effect

was observed ($F(1,11)=5.448, p<0.05$) suggesting that arousal increased when comparing Low-

arousal and High-arousal categories (High-arousal showed a greater P300 amplitude), but only when the stimulus was in the form of a word. Channel 2 within the Group groupings categorization revealed only a main effect of stimulus ($F(1,11)=13.535, p<0.05$), but not of valence ($F(1,11)=0.392, p=0.545$) or an interaction effect ($F(1,11)=2.380, p=0.154$). In Channel 2, the main effect of stimulus indicated the same trend in which pictures consistently elicited higher amplitudes than words (**Figure 7**).

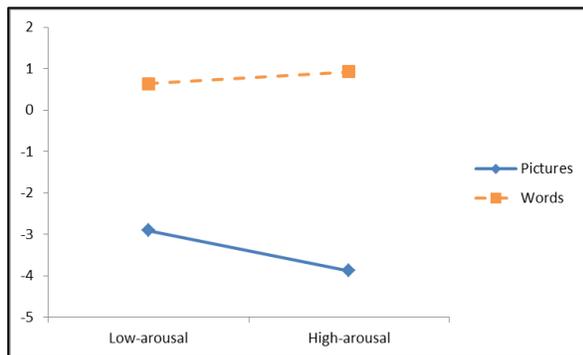


Figure 7 shows the P300 output for Channel 2 within the Group groupings classification. This graph shows a main effect of stimuli where pictures show greater brain activity than words. This graph exemplifies the findings found in other areas of P300 results as well.

The Individual groupings classification yielded similar results across both Channels. In both Channel 1 and 2, a main effect was observed for stimuli (Channel 1: $F(1,11)=31.144, p<0.001$; Channel 2: $F(1,11)=14.686, p<0.05$). Similarly to results of main effect of stimulus found in the Group groupings system, both channels in the Individual groupings revealed consistently higher amplitudes associated with pictures over words. Neither Channel showed a significant main effect of valence (Channel 1: $F(1,11)=0.380, p=0.551$; Channel 2: $F(1,11)=0.003, p=0.955$) or a significant interaction effect (Channel 1: $F(1,11)=1.346, p=0.273$; Channel 2: $F(1,11)=0.091, p=0.769$).

LPP Analysis

After running the same Repeated Measures ANOVA analysis on the LPP component, relatively similar results ensued. No significance was revealed in the Group groupings classification in Channel 1 for main effect of stimulus ($F(1,11)=4.337, p=0.064$), main effect of

valence ($F(1,11)=4.197, p=0.068$), or an interaction effect ($F(1,11)=1.518, p=0.246$). Channel 2 within Group groupings did indicate a significant main effect of stimulus ($F(1,11)=55.408, p<0.001$) in which pictures were consistently more arousing than words. However, no main effect of valence ($F(1,11)=2.948, p=0.117$) or interaction effect ($F(1,11)=0.007, p=0.935$) were observed (**Figure 8**).

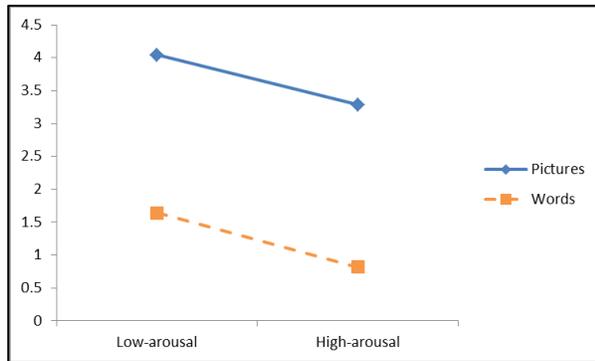


Figure 8 shows the LPP output for Channel 2 within the Group groupings classification. Similarly to Figure 7, this graph shows a main effect of stimuli in that pictures consistently elicit higher amplitudes than words. This graph also serves as a representation of the findings found in other areas of LPP results as well.

Within the Individual groups classification for the LPP component, Channels 1 and 2 both displayed a significant main effect of stimulus (Channel 1: $F(1,11)=8.573, p<0.05$; Channel 2: $F(1,11)=52.919, p<0.001$). Across both channels, data showed that pictures revealed higher amplitudes than words. There was no main effect of valence observed in either Channel (Channel 1: ($F(1,11)=1.956, p=0.192$; Channel 2: $F(1,11)=0.612, p=0.452$), nor was there an interaction effect (Channel 1: $F(1,11)=0.018, p=0.895$; Channel 2: ($F(1,11)=2.207, p=0.168$).

Discussion

After running analysis using the new subgroups of High-arousal Stimuli and Low-arousal Stimuli, it is evident there is no significant difference between the individually assigned subjective ratings of stimuli by participants, regardless of the Channel analyzed (Channel 1 or Channel 2), the classification system used (Group groupings or Individual groupings), or the component analyzed (P300 or LPP). Regardless of whether a stimulus was in the form of a picture or word, and regardless of whether it was identified as extremely negative or extremely

positive, we expected to see higher P300 and LPP amplitudes in the High-arousal Stimuli category when compared to those of the Low-arousal Stimuli. The overall lack of main effects of valence was therefore not in line with the predictions established, but there could be a number of reasons explaining this.

Once the original group of participants was narrowed down to the final 11, those 11 participants were included in every level of analysis. However, due to the nature of their subjective ratings, some of these participants may have had as few as 10 or as many as 170 valid trials with which to analyze. Beyond this initial check for validity, no single participant ended up with 100% of valid ERPs that corresponded with their ratings; the highest percentage of ERPs among participants was 94.55% while the lowest percentage was as low as 25.00%. The fewer valid ERPs each participant had to draw from, the less reliable their data was. It is worth noting, however, that the lack of significant main effects and interaction effects in the Repeated Measures ANOVA of percent valid EEG discounts this possibility as serving as a confound. Nor were there significant results obtained through the analysis of the difference between high and low ratings across participants. These two analyses served as checks to ensure that though they may have played a minor role the lack of significant results, they were not significant enough to affect the statistics.

While there was one instance among P300 analysis where we saw an interaction effect, in the Group groupings classification of Channel 1, this significance could potentially be attributed to the possibility of a Type I error. Type I errors emerge when we inappropriately reject the hypothesis which is in fact true. Having an alpha value of 0.05 means that there is a 5% chance that the significance revealed is not truly significant. Though it would be impossible to determine whether or not this is the case, it is mathematically very likely in light of the amount

of analysis performed. Since the analysis involved 8 categories (classification: group groupings, individual groupings; Channel: 1, 2; component: P300, LPP) across 2 stimulus types (pictures, words), there were 16 separate analyses happening. The chances that 1 finding in 16 was statistically inaccurate approximates 5%. Since this finding doesn't align with our predictions (if anything, pictures should be changing, not words), we might be inclined to attribute it to a Type I error rather than something that simply cannot be explained otherwise (**Figure 9**).

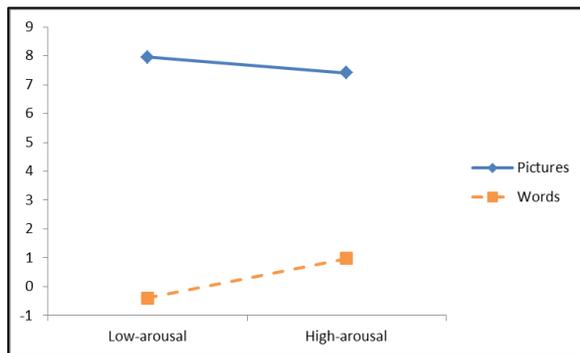


Figure 9 shows the P300 output for Channel 1 within the Group groupings classification. While an interaction effect is observed (in which arousal level affects higher P300 amplitudes, but only for words), we have reason to believe that this significant finding may be due to a Type I error. The fact that the interaction effect is present only in words, but not pictures, is counterintuitive and cannot be explained by our predictions. Further, the likelihood that an error may have emerged is convincing (6.25%).

Further, while we failed to observe any main effects of valence, the fact that we consistently came across main effects of stimuli in the P300 and LPP analysis is reassuring. If we hadn't seen these significant results, we would have reason to worry. What these main effects indicate is a substantial difference between the way our brains process the two types of stimuli involved—pictures and words. Intuitively, this should be the case. Pictures and words are intrinsically different (i.e. pictures were much larger, had more details and more aspects to focus on and process) and therefore would be expected to reveal a difference in the physiological responses emitted via brain activity.

Based on these results, it would be possible to conclude that input that is self-reported as being very positive or very negative actually do not necessarily correspond directly with similar physiological responses, which we would expect to be evidenced by increased amplitudes of the P300 and/or LPP components. Similarly, we can conclude that stimuli that are perceived as

subjectively more arousing are not guaranteed to be physiologically more arousing as well. These results would seem to imply that our perceptions and impressions do not quite mimic underlying functions of the brain. Due to the lack of significance, we cannot reject the null hypotheses and are resigned to accept that either the research design was flawed in some fundamental way or that the trends we expected to see are due to chance.

While in many conditions, Repeated Measures ANOVA revealed that the arousal evidenced by P300 and LPP amplitudes corresponding with picture and word stimuli increased from Low-arousal Stimuli to High-arousal Stimuli, there are also many instances where this effect is reversed/counter to our expectations. These results may indicate a selective deeper semantic processing associated with more emotionally-charged words.

While our design was conceptually similar to that of Schupp, et al (2007) in terms of varying the valence of visual images, the categories of their images were more extreme. Rather than including general positive, neutral, and negative stimuli, their images consisted more specifically of erotica, neutral images, and mutilation. Although their categories do not coincide directly with those of the original study, the essence of them imitates ours on a more extreme level. We expected similar results based on these intrinsic similarities, but failed to find them.

Despite the lack of statistical support for my original hypotheses, the most constructive aspect of designing and carrying out this analysis was the opportunity to explore, to ask questions, and to design a way in which my questions could be answered. The logistics of the approach helped me to formulate, consider, and reconsider my ideas and how I could test them. The ability to devise a question, assess literature that might touch on it, and create the methodology with which I could systematically and scientifically learn provided a larger picture of the research process.

References

- Citron FMM (2012) Neural correlates of written emotion word processing: A review of recent electrophysiological and hemodynamic neuroimaging studies, *Brain Lang* 122: 211-226.
- Hajcak G, Olvet DM (2008) The persistence of attention to emotion: Brain potentials during and after picture presentation, *Emotion* 8:250-255.
- Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: A module in human extrastriate cortex specialized for face perception, *J Neurosci* 17:4302-4311.
- Kensinger EA, Schacter DL (2006) Processing emotional pictures and words: Effects of valence and arousal, *Cogn Affective Behav Ne* 6:110-126.
- Kissler J, Herbert C, Peyk P, Junghöfer M (2007) Buzzword: Early cortical responses to emotional words during reading, *Psychol Sci* 18:475-480.
- Kuriki S, Takeuchi F, Hirata Y (1998) Neural processing of words in the human extrastriate visual cortex, *Cognitive Brain Res* 6:193-203.
- Mallan KM, Lipp OV (2011) The relationship between self-reported animal fear and ERP modulation: Evidence for enhanced processing and fear of harmless invertebrates in snake- and spider-fearful individuals, *Motiv Emotion* 35:474-483.
- Nechvatal JM, Lyons DM (2013) Coping changes the brain, *Front Behav Neurosci* 7:1-8.
- Schienle A, Schäfer A, Naumann E (2008) Event-related brain potentials of spider phobics to disorder-relevant, generally disgust- and fear-inducing pictures, *J Psychophysiol* 22:5-13.
- Schupp HT, Junghöfer M, Weike AI, Hamm AO (2004) The selective processing of briefly presented affective pictures: An ERP analysis, *Psychophysiology* 41:441-449.
- Schupp HT, Stockburger J, Codispoti M, Junghöfer M, Weike AI, Hamm AO (2007) Selective visual attention to emotion, *J Neurosci* 27:1082-1089.

Weinberg A, Hajcak G (2010) Beyond good and evil: The time-course of neural activity elicited by specific picture content, *Emotion* 10:767-782.

SAMPLE