Processing Emotion and Taboo in a Native vs. a Second Language: An ERP Study

Allison R. Fogel

Tufts University

Abstract

Bilinguals frequently report that words in a second language (L2) do not seem to have as much emotional impact as words in a native language (L1) and that it is easier to swear in L2. This intuition has been corroborated experimentally, most reliably by studies that have found greater skin conductance responses (SCRs) in response to emotional phrases and taboo words in L1 than in L2, but the processing differences underlying this disparity remain unclear. In monolinguals, the late positive component (LPC) has been shown to be responsive to a word's level of arousal (intensity) but not to its valence (pleasantness). The present study investigated how these dual aspects of a word's emotionality are processed in L2. In Experiment 1, participants performed a semantic categorization task as they read single words that were crossed in levels of valence and arousal. As predicted, highly arousing words evoked a larger LPC than low arousal words in native English speakers but not in L2 English speakers, and no main effect of valence was observed in either group. Experiment 2 added taboo words to the same design. As with the arousal effects seen in Experiment 1, taboo words elicited a robust LPC compared to neutral words in native speakers, while the taboo effect was restricted in distribution in L2 speakers. However, no arousal effects were observed to highly arousing nontaboo words in either group when taboo words were present in the paradigm. These results suggest that L2 words are experienced as less arousing than L1 words and that arousal is relative to the surrounding context.

Processing Emotion and Taboo in a Native vs. a Second Language: An ERP Study

There is a common perception that words in one's second language are not "felt as deeply" as a native language. Anecdotal evidence for this phenomenon is abundant; bilinguals report preferring to use their first language (L1) to express strong emotions, as opposed to their second language (L2) that is regarded as being more distant and impersonal (Dewaele & Pavlenko, 2002; Pavlenko, 2007). It has been reported that bilinguals find it easier to discuss embarrassing topics in their L2 and may even code-switch to L2 in order to do so (Bond & Lai, 1986). Similarly, bilinguals frequently report that it is easier to swear in a non-native language, as swear and taboo words do not have as much impact (Dewaele, 2004). Accordingly, taboo words have been reported to arouse more anxiety in L1 (Gonzalez-Reigosa, 1976). As a growing proportion of the world's population becomes bi- and multilingual, it is increasingly relevant to understand the dynamics of second language use.

It has been well-established that emotionally evocative words are capable of capturing attention. Affective words' emotionality is most frequently quantified along the two dimensions of hedonic valence (pleasantness) and arousal (intensity) (Fischler & Bradley, 2006). High valence words are pleasant, while low valence words are unpleasant; high arousal words are exciting and salient, while low arousal words are unexciting or dull. According to the model of affective space described by Bradley and Lang (1999) in their Affective Norms for English Words (ANEW) these two variables are not inexorably linked, that is, it is possible for highly pleasant or unpleasant words to be either high or low in arousal. However, as seen in Figure 1, words at the extremes of the valence measure tend to be highly arousing, and it is worth noting that high valence (pleasant) words are generally more widely distributed in levels of arousal than low valence (unpleasant) words.

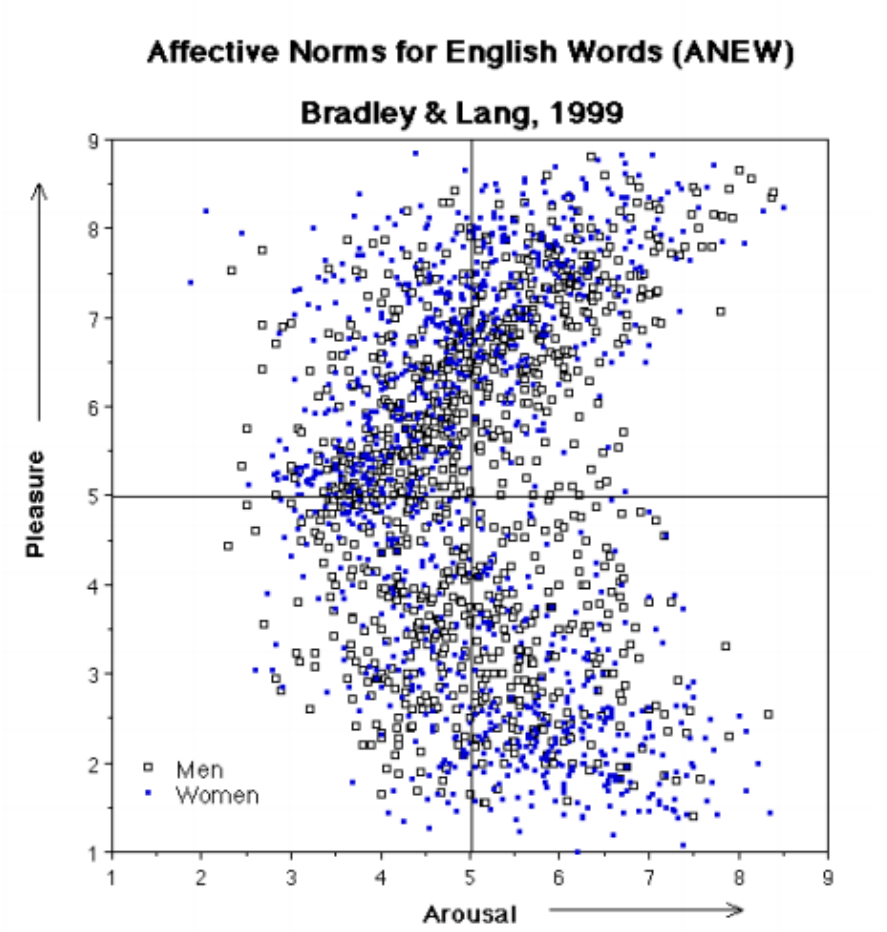

Figure 1. Distribution of the Affective Norms for English Words (ANEW) corpus (Bradley & Lang, 1999)

Taboo words represent a special category of highly emotional language that includes profanity, sexual terms and slurs. Ratings studies have found that taboo words are more arousing than nontaboo words, including emotional words, but that they are not of more extreme valence. While their overall valence rating is comparable to that of unpleasant emotional words, taboo words are processed more automatically and cannot be considered as being part of the same category (Janschewitz, 2008). There is evidence from multiple domains that taboo words are extremely salient due to their high levels of emotional arousal. Taboo words are recalled more frequently than neutral words regardless of task instructions, and they evoke stronger skin conductance responses (SCR, a psychophysiological measure of autonomic arousal) than

nontaboo emotional words (Jay, Caldwell-Harris, & King, 2008). In the taboo Stroop task, where participants are instructed to name the color of taboo and neutral words, color naming times are longer for taboo than for neutral words, and the colors previously associated with these words are recognized more accurately (MacKay et al., 2004). Arnell, Killman, & Fijavz (2007) included taboo words in an RSVP (Rapid Serial Visual Presentation) task, in which words are presented one at a time at a high rate of speed. It has been shown in this task that highly arousing words receive preferential attentional processing and inhibit identification of subsequent words, an effect known as an attentional blink. When taboo and sexual words served as distracters before a color word target, participants' target accuracy rates were lower than when pleasant, unpleasant or neutral words were used.

A wide range of techniques have been used to investigate the processing of emotional and taboo words in L2 experimentally, often yielding conflicting results. In an influential study, Anooshian and Hertel (1994) tested the emotional memory advantage in late English-Spanish and Spanish-English bilinguals. More emotional than neutral words were recalled when stimuli had been presented in bilinguals' native languages, an effect that was not present in their L2. This study provided important objective evidence for the subjective intuition that L1 words are more emotional. However, some more recent investigations have failed to replicate the main finding; Ferré, Garcia, Fraga, Sanchez-Casas, & Molero (2010) tested both early and late bilinguals and found that highly arousing pleasant and unpleasant words showed the same memory advantage over neutral words in L1 and L2. Ayçiçeği and Harris (2004) included taboo words in their design, using a population of late Turkish-English bilinguals. Results indicated that the emotion memory advantage was actually larger in English (L2), and taboo words showed

an advantage in recall and recognition in both languages, contradicting Anooshian and Hertel's (1994) conclusion.

In another line of investigation, researchers have compared bilinguals' performance on the emotional Stroop task in their different languages. In this task, subjects are shown emotional and neutral words and asked to identify the color of the text; color naming response latencies have been shown to be slower in response to unpleasant and taboo words (MacKay et al., 2004). This task was used by Sutton, Altarriba, Gianico, & Basnight-Brown (2007) with a population of early Spanish-English bilinguals, who found a larger interference effect of unpleasant words in English (L2); the authors note that this is somewhat expected, given that their participants were more proficient and likely dominant in L2. Similarly, Eilola, Havelka, & Sharma (2007) presented a Stroop task that contained neutral, pleasant, unpleasant, and taboo words to a group of native Finnish speakers who were late but proficient learners of English. It was found that response latencies were longer to unpleasant and taboo words than neutral words in both Finnish (L1) and English (L2), and the magnitude of the interference was the same in both languages. The authors propose that this may indicate that level of proficiency in L2 is the crucial factor modulating these effects, as their late bilingual subjects would be expected to show a difference if age of acquisition was influential, but also suggest that the fact that participants were told to ignore the meanings of words to complete the task could be responsible for the lack of differential effects.

The RSVP paradigm is an additional task that reflects automatic emotional activation. As discussed above, it has been found that the presentation of a taboo word results in an attentional blink that reduces accuracy of reporting the subsequent word (Arnell et al., 2007). Colbeck and Bowers (2012) presented late Chinese-English bilinguals and native English speakers with a

RSVP task in which taboo, neutral and nonwords served as distracters before color word targets. Target words preceded by taboo words were identified more poorly by native English speakers than by native Chinese (L2 English) speakers, indicating that the attentional blink, while robust in L1 speakers, was reduced in L2 speakers. This suggests that, at least for late learners of a language, the highly arousing properties of taboo words are less salient in L2 than in a native language.

Studies measuring skin conductance responses (SCRs) have provided some of the most convincing evidence of diminished responses to emotional words in a second language. In the study conducted by Harris, Ayçiçeği, & Gleason (2003), late Turkish-English bilinguals who had acquired English after age 12 either heard or read neutral, pleasant, unpleasant, and taboo words and childhood reprimands. The strongest responses were elicited by taboo words in either language. For all types of words, larger SCRs were elicited by Turkish (L1) than by English (L2) items, with the largest differences occurring for reprimands in either mode of presentation and for taboo words in the auditory modality. These results were also replicated in an L1 context, ensuring that the effects were not simply a result of hearing L1 in an atypical environment (Caldwell-Harris & Ayçiçeği-Dinn, 2009). Harris (2004) used a similar procedure with Spanish-English bilinguals who had learned English in either early childhood (immersed by age 7) or middle childhood (first exposed in a classroom setting at age 8-12). Childhood reprimands elicited larger SCRs in L1 than in L2 for late learners only; in general, differences between languages were small. These varying findings from studies where participants had learned English at different ages might seem to implicate age of acquisition of a language as the main predictor of its emotionality, but the authors note that proficiency in a language and the age at which it was learned were generally covariant in their participants. As an explanation for their

results, they propose an "emotional contexts of learning" theory, according to which languages are experienced as emotional to the extent to which they were learned and are used in emotionally charged contexts (Harris, Gleason, & Ayçiçeği, 2005).

The existing research has clearly not arrived at a consensus on the subject of the emotionality of bilinguals' languages. It is difficult to draw broad conclusions from these studies, as they have all varied on several measures as basic as the task and the nature of the stimuli that were used. Even the criteria by which participants are classified as early or late bilinguals are inconsistent between researchers. Though there is reason to believe that the context in which a language was learned may influence its emotionality (Dewaele, 2004), studies have included both classroom and immersive learners. Most research designs have presented words in L1 and L2 to participants who share the same two languages. While this method ensures consistency between groups, it is also susceptible to the influence of culturally specific attitudes; concepts of emotion and taboo differ between cultures. A word may be used much more casually in one language than its translation in another, and many participants acquired their L2 by moving to a new country entirely. Additionally, some tasks may be better suited to capturing differences than others. Many memory and Stroop tasks have found equal emotion effects in bilinguals' languages, but this does not necessarily indicate that no differences exist; these methods may simply not be sensitive enough to reflect the variance between languages.

Event-related potentials (ERPs) are measured by recording changes in electrophysiological activity on the scalp in response to the presentation of a stimulus. Because of their high temporal resolution, ERPs are well-suited to studying to studying the time-course of word processing, including the presumably subtle effects of emotion. ERP studies of single emotional words have most reliably found a late positive component (LPC) consisting of a

greater positivity elicited in response to emotional than to neutral words approximately 500 ms after stimulus presentation. However, it has so far been unclear which aspects of an emotional word trigger the LPC. There are some reports of differential effects for pleasant and unpleasant words, though, as noted by Fischler & Bradley in their 2006 review, no consistent pattern has been observed. Some studies have found a greater positivity in response to pleasant words (Kissler, Assadollahi, & Herbert, 2006; Kissler, Herbert, Winkler, & Junghofer, 2009; Herbert, Junghofer, & Kissler, 2008), while others have shown that pleasant and unpleasant words result in the same response (Fischler & Bradley, 2006). This variance may be due in part to task effects. In order to address this issue, Fischler and Bradley (2006) conducted a series of experiments in which they varied the nature of the tasks participants completed while reading the same set of stimuli. When participants judged words' pleasantness or emotionality, pleasant and unpleasant words elicited an equally robust effect on the LPC. Pleasant and unpleasant words that were read with no task in mind resulted in effects that were equal but less robust, though still significant. However, only unpleasant responses elicited a significant LPC effect when participants performed a semantic categorization task, and no effects were observed with a lexical decision task. In addition to this complicated set of results, another issue with the studies discussed above is that most used only highly arousing pleasant, highly arousing unpleasant and neutral words as stimuli, confounding the variables of valence and arousal. Delaney-Busch, Haime, Wilkie, & Kuperberg (2011) were the first to investigate these variables separately, including negative, neutral and positive words with high and low levels of arousal in their design. It was found that the LPC was triggered only in response to arousal; valence did not have an effect on the amplitude of this component.

Only one published study has previously attempted to use ERPs to investigate the relative emotionality of words in bilinguals' languages. Conrad, Recio, & Jacobs (2011) presented pleasant, unpleasant, and neutral words in both German and Spanish to two groups of late German-Spanish and Spanish-German bilinguals. For native German speakers, L2 emotional words produced effects similar to those of L1 words, while only positive L2 words resulted in a significant effect in native Spanish speakers; additionally, there was a latency shift between languages, with "valence effects" delayed by approximately 50 ms in L2. The authors conclude that emotional processing is qualitatively the same but delayed in L2. However, this interpretation is questionable in light of several potential confounds in the experimental design. The two different groups of participants varied in their L2 proficiency and age of language acquisition. Most significantly, emotional stimuli were one again not controlled for arousal either within or between languages, despite the fact that arousal is known to be the main inducer of the LPC.

As previous experiments have provided conflicting results, it is apparent that a clearer examination of the impact of emotional words in L2 could provide more useful information. In Experiment 1, we aimed to investigate how valence and arousal may be processed differently in L2 with ERPs in a more controlled manner than Conrad et al. (2011). By using a design similar to that of Delaney-Busch et al. (2011), it is possible to examine the relative contributions of these two aspects of emotionality. A semantic categorization task that did not require participants to attend specifically to the emotional aspects of words was used. This task is well adapted to emotion research, as it has been shown that task demands can affect ERP responses to emotional words (Fischler & Bradley, 2006; Delaney-Busch et al., 2011). As in natural reading, participants

completing this task consider the semantic content of words with no particular characteristic in mind.

We predicted that highly arousing words would evoke a more positive LPC than low arousal words and that valence would not affect the LPC in native English speakers, as has previously been found (Delaney-Busch et al., 2011). It was hypothesized that this arousal effect would be diminished in participants who learned English as their second language, reflected as a smaller or no difference between the effects of high and low arousal words on the LPC than in native speakers.

Experiment 2 sought to extend our investigation of the emotionality of bilinguals' languages to the domain of taboo words. To our knowledge, no previous ERP study has examined the processing of taboo words; as it has been suggested that taboo words are highly emotional due to their high levels of arousal, we expected these words to evoke an LPC resembling that seen in response to high arousal words. In order to be able to make a valid comparison between taboo and highly arousing nontaboo words in the same population, taboo words were added to the Experiment 1 paradigm. As in the pattern seen with highly arousing words in Experiment 1, it was predicted that the taboo word effect would be attenuated in L2 speakers.

Experiment 1

Methods

Participants

32 Tufts University students (16 female, mean age = 20.7, *SD* = 1.5) were recruited and paid for their participation. All participants were right handed, had normal or corrected-to-

normal vision and reported no history of traumatic head injury. Sixteen participants had learned English as their L2, beginning at a mean age of 5.3 years (*SD* = 2.55). Participants in this group spoke ten different languages (Spanish, Farsi, Russian, Vietnamese, Hebrew, Arabic, Lai, Chinese, Cantonese, and Korean) as their L1. Of the 16 L2 English speakers, 11 were early bilinguals who learned English at age 1-6 and five were late bilinguals who learned English at age 6-10. The remaining 16 participants were native English speakers who had not been exposed to any other language before the age of 6 and had not had any language learning experience in the past two years.

Participants rated their abilities to read, speak and comprehend their L1 and L2 as well as how frequently they read and communicated in each language on a 7-point Likert scale (1 = *unable* to 7 = *expert*). Native English speakers' average level of English ability was 6.9 (*SD* = 0.40), higher than that of participants who learned English as a L2 who rated their ability as 6.6 (*SD* = 0.63) ($t(126) = 2.68$, $p < 0.01$). Both groups reported that English was the language they used most frequently; on a 7-point Likert scale (1 = *rarely*; 7 = *very frequently*), L1 English speakers rated their frequency of use as 6.9 (*SD* = 0.39) and L2 English speakers reported an English use frequency of 6.7 (*SD* = 1.00) ($t(62) = 1.31$, $p = 0.20$).

Stimuli

The stimuli used in this study were 240 English words. These items were a subset of the stimuli used by Delaney-Busch et al. (2011), who developed a corpus of 466 words that were rated on their levels of emotional valence, arousal and concreteness by online participants. Each item used in the present experiment was between three and thirteen letters, and the mean length of all items was 6.94 letters (*SD* = 1.81). The average frequency of the words as occurring in the

Hyperspace Analogue to Language (HAL) frequency norms was 102.38 occurrences per million (*SD* = 268.11). Levels of valence and arousal were crossed to create groups of 40 words with high valence and high arousal (VA), 40 with high valence and low arousal (Va), 40 with low valence and high arousal (vA), and 40 with low valence and low arousal (va). (See Table 1 for characteristics of these items; see Appendix A for complete list of critical items.) High arousal items significantly differed from low arousal items on the measure of arousal ($t(158) = 29.3$, $p < 0.01$) and did not significantly differ in valence, length or frequency (all ps > 0.15), though they did differ in concreteness with low arousal items tending to be more concrete. High valence items significantly differed from low valence items on the measure of valence ($t(158) = 56.3$, $p < 0.01$) and did not significantly differ in arousal, concreteness, length or frequency (all ps > 0.15). In order to make the experimental manipulation less evident, 80 filler items were included. Of these, 40 were neutral items with medium ratings of both valence and arousal and 40 had relatively neutral valence and arousal ratings.

	Valence	Arousal	Concreteness	Length	Frequency
VA	5.71 *(0.32)*	5.06 (*0.38)*	3.54 *(0.93)*	7.08 (*1.76)*	76.53 (*132.53)*
Va	5.43 *(0.39)*	3.11 *(0.39)*	4.07 *(1.23)*	6.93 *(2.09)*	186.67 *(486.02)*
vA	1.92 *(0.39)*	4.97 *(0.45)*	3.73 *(1.02)*	7.05 *(1.81)*	53.75 *(99.59)*
va	2.28 *(0.32)*	3.17 *(0.40)*	4.02 *(1.03)*	6.35 *(2.09)*	106.43 *(237.98)*
Neutral	3.62 *(0.74)*	4.10 *(0.09)*	3.85 *(0.77)*	7.20 *(1.54)*	95.04 *(178.45)*

Table 1. Mean valence, arousal, concreteness, length, and frequency (occurrences per million) values of words used in the experiment.

Forty probe items of four to eleven letters in the semantic category of "animal names" (such as *cat* or *dog*) were included in the presented lists (14% of all items). Each participant saw every experimental and probe item in one of six randomly sorted orders.

Procedure

Participants were seated in a comfortable chair with their eyes approximately 1.5 meters away from a monitor. The word stimuli were presented as white letters centered vertically and horizontally on a black background. The maximum height and width of the stimuli were such that no saccades would be required during reading of the single word stimuli. Participants completed a go/no-go semantic categorization task in which they were instructed to press one button on a game pad to words that were animal names. Each trial consisted of a fixation cross lasting 1800 ms, a blank screen for 500 ms, an item appearing for 450 ms, and then a blank screen for 700 ms (see Figure 2 for an example trial). A blink stimulus ["(- -)"] lasting 2500 ms appeared after every 10-15 items, followed by a blank screen for 500 ms before the presentation of the next trial.

After electrode placement, instructions for the experimental task were given and a short practice list was presented to familiarize participants with the task. The experimental section typically took about 15 minutes to complete and contained two pauses of a length determined by the participant.

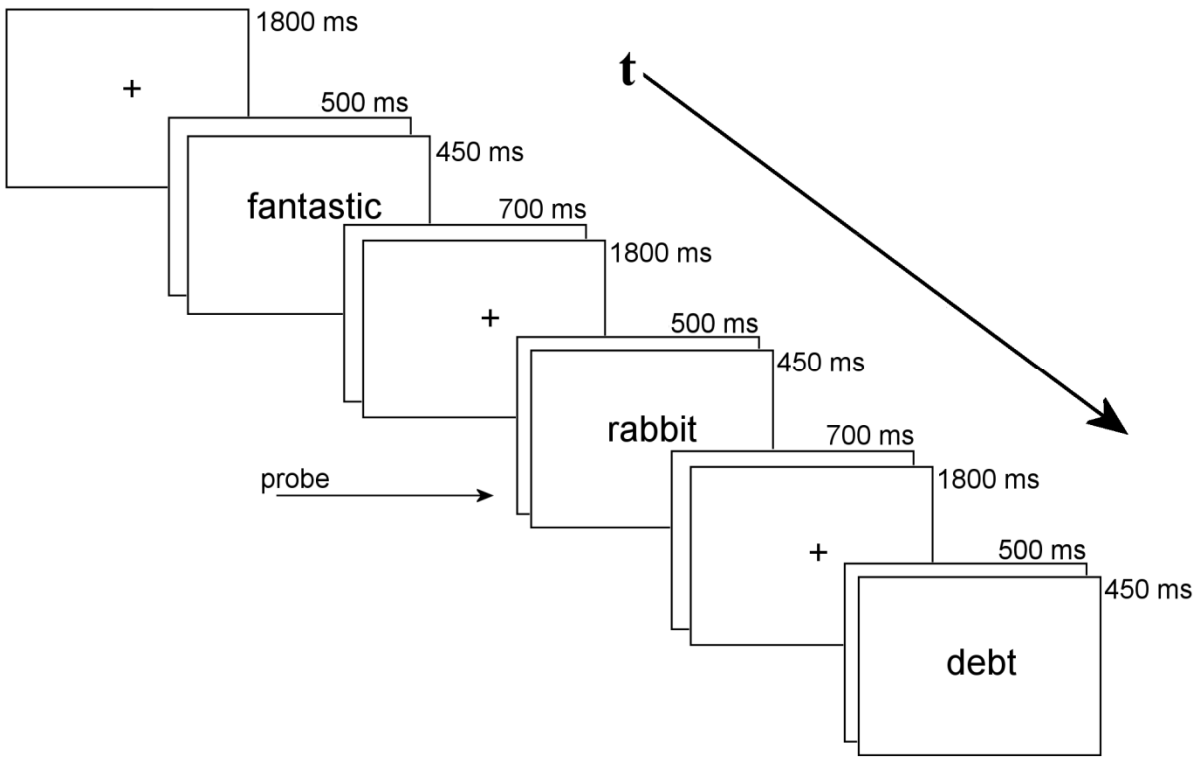

Figure 2. A typical experimental trial.

EEG recording

Participants were tested while seated in a comfortable chair in a darkened sound attenuated room. The electroencephalogram (EEG) was recorded from 29 tin electrodes embedded in an elastic cap (Electro-Cap International). Two additional electrodes were attached below the left eye (LE) and to the right of the right eye (VE) to monitor blinks and eye movements. All electrodes were referenced to an electrode placed over the left mastoid bone (A1), and an electrode on the right mastoid bone (A2) was used to monitor for differential mastoid activity; none was observed. Electrode impedances were maintained below 5 kΩ for the 29 head electrodes, below 10 kΩ for the eye electrodes, and under 2 kΩ for the mastoid electrodes. The EEG data was amplified using an SA Instruments Bio-amplifier system at a bandpass of 0.01-40 Hz and digitized at a rate of 200 Hz throughout the experiment.

Data analysis

Averaged ERPs were formed off-line from trials free of ocular and muscular artifact (9% of all trials were rejected due to artifact) and were lowpass filtered at 15 Hz. A subset of 12 of the 29 scalp electrodes was chosen for analysis (see Figure 3). Mean amplitude was measured in the epochs of 150-250, 300-500, and 500-700 ms. Average waveforms were created for two levels of GROUP (L1 and L2) two levels of VALENCE (High and Low), two levels of AROUSAL (High and Low), three levels of LATERALITY, and four levels of POSTERIOR. Repeated analyses of variance (ANOVAs) were performed on the data in each epoch. The Geisser and Greenhouse (1959) correction was applied to all repeated measures with more than one degree of freedom in the numerator.

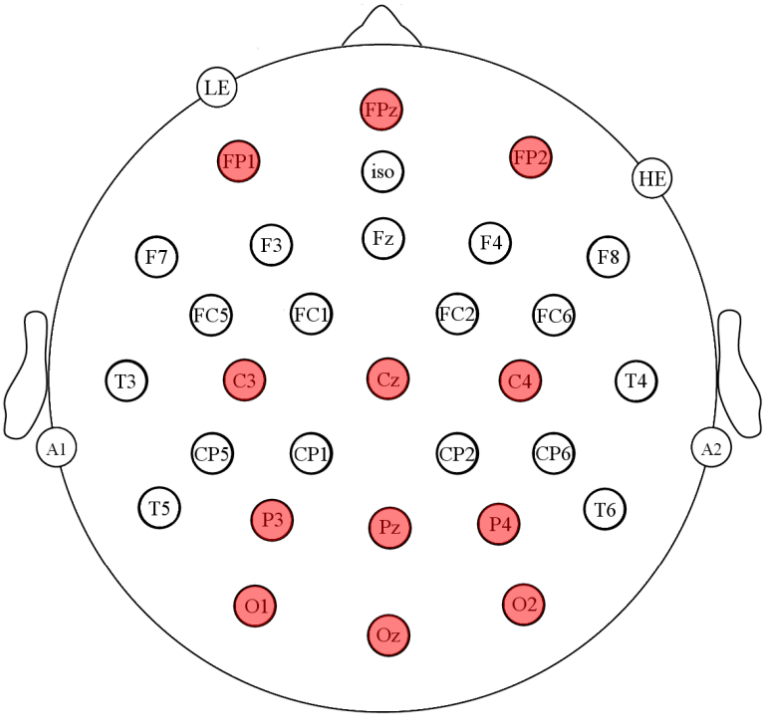

Figure 3. Electrode montage showing the locations of 29 scalp electrodes. Electrodes highlighted in red represent electrode sites used in data analysis (FP1, FPz, FP2, C3, Cz, C4, P3, Pz, P4, O1, Oz, O2).

Results

Visual Inspection of ERPs

Grand-average ERP waveforms elicited by high and low arousal words for each participant group are shown in Figures 4a and 4b. Figure 5 shows voltage maps created by subtracting the ERPs for low arousal items from the ERPs for high arousal items for each group. ERP waveforms elicited by high and low valence words for each group are shown in Figure 6a and 6b. Voltage maps created by subtracting ERPs to low arousal items from ERPs to high arousal items within each level of valence are shown in Figure 7. ERPs at electrode F3 resulted in an average waveform that stood out from all nearby electrodes; it is improbable that this could have been elicited by cortical activity and was more likely to have been a result of experimenter error or equipment failure. Therefore, this electrode was eliminated from all voltage maps and analyses. As can be seen in Figures 4a and 4b, the first visible component was a small negativity between 90 and 150 ms after stimulus onset (N1), followed by a positivity peaking around 150-200 ms (P2). In the L1 group, there are differences at anterior sites between low and high arousal words on the negativity between 300 and 500 ms (N400), with an attenuation of the N400 for high arousal compared to low arousal words. The positivity between 500 and 700 ms (LPC) is more pronounced for high arousal than for low arousal words, with the largest differences occurring at centro-posterior sites.

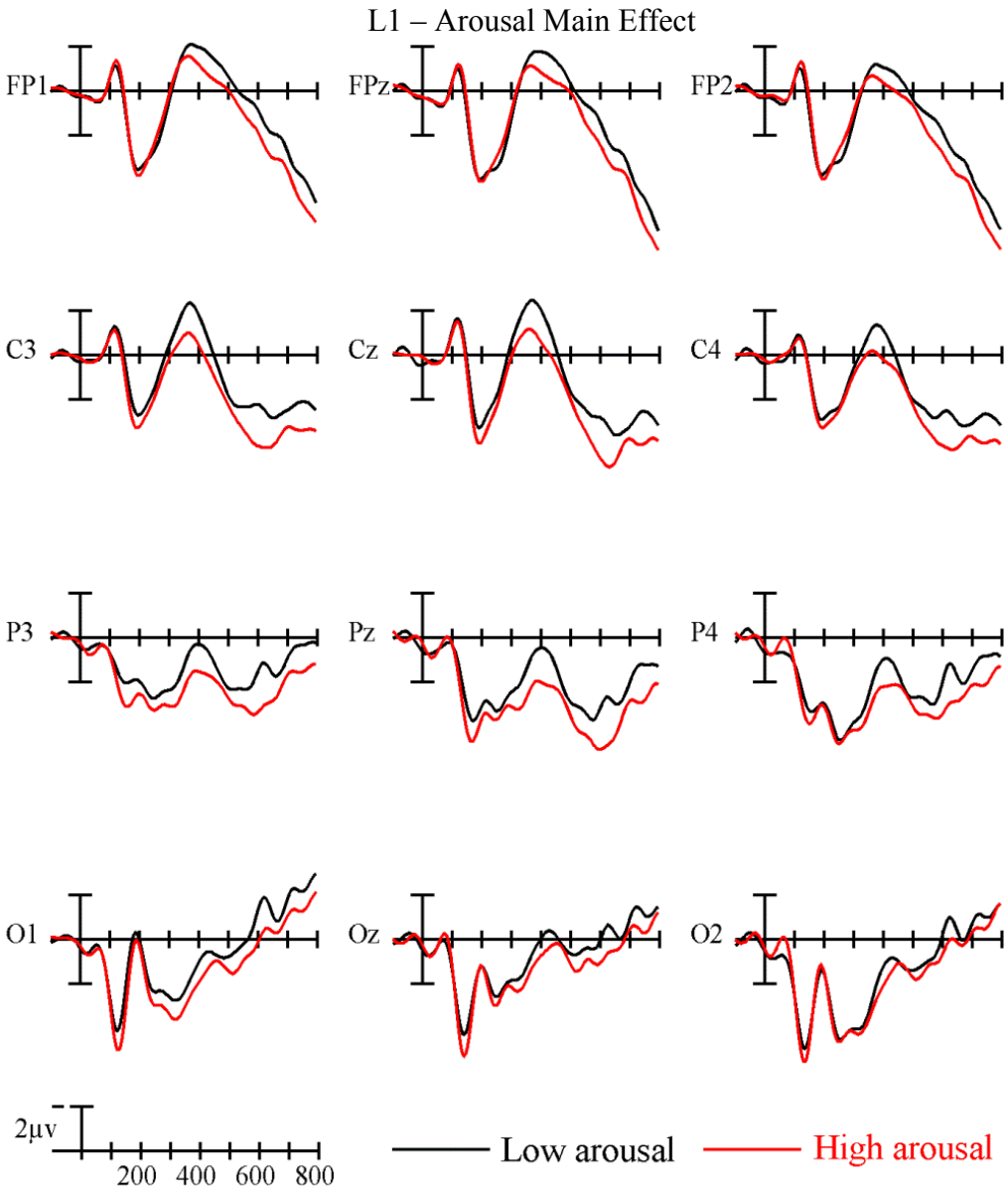

Figure 4a. ERPs to low and high arousal items in native English speakers.

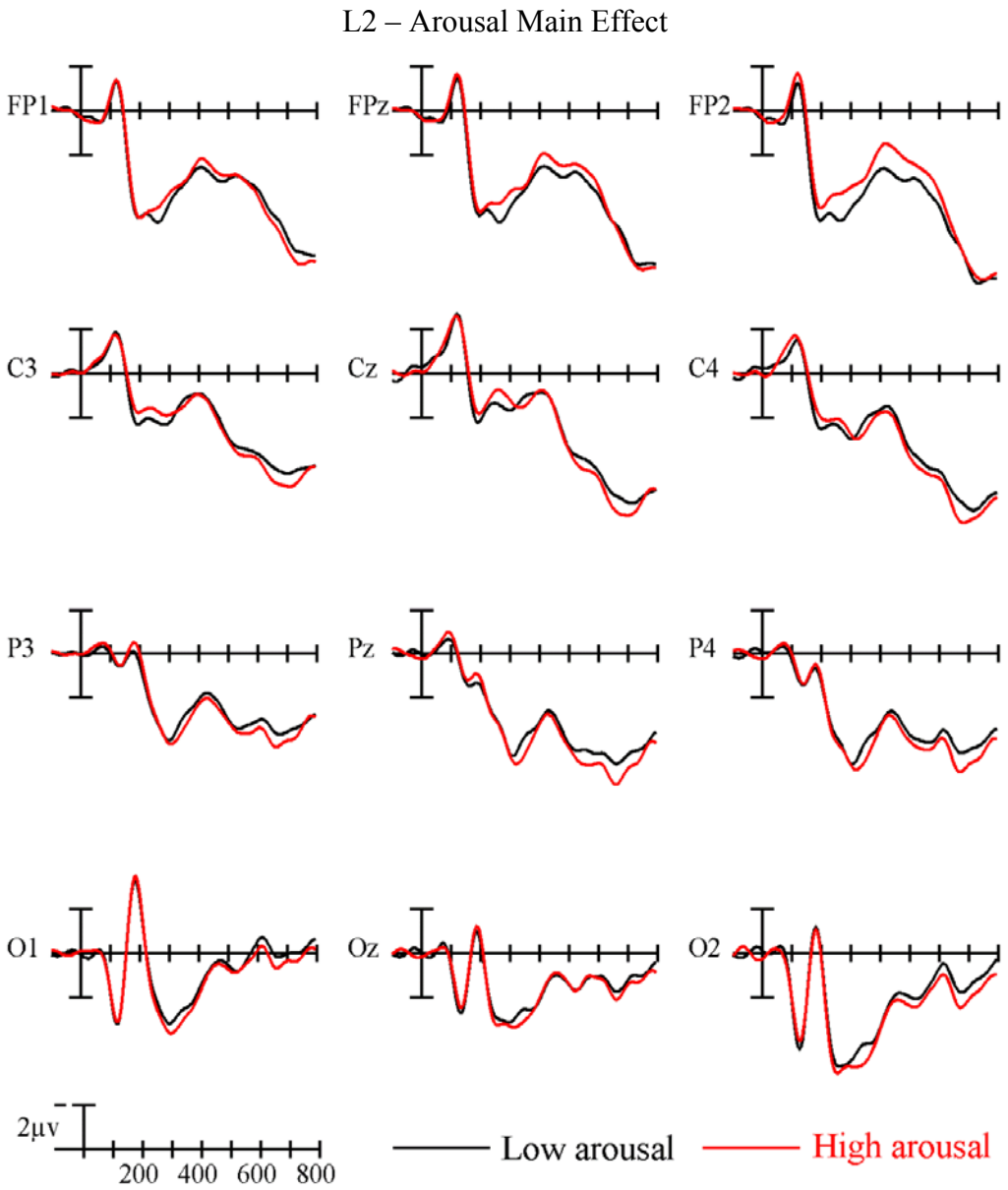

Figure 4b. ERPs to low and high arousal items in L2 English speakers.

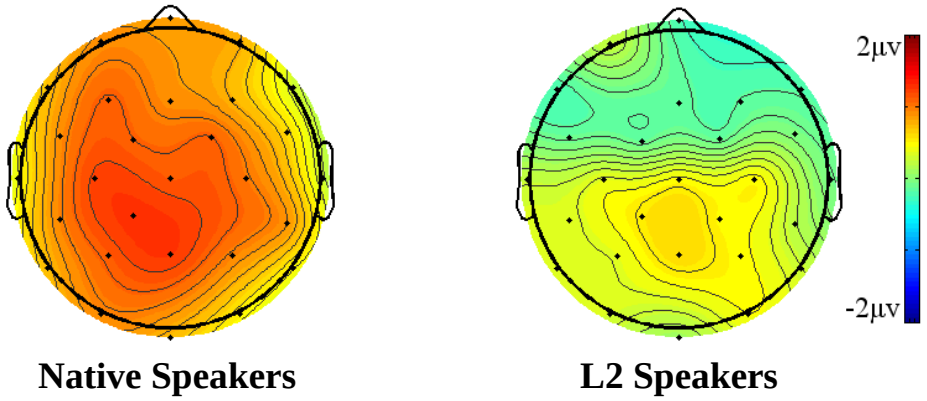

Figure 5. Voltage maps of low arousal subtracted from high arousal words from 500-700 ms.

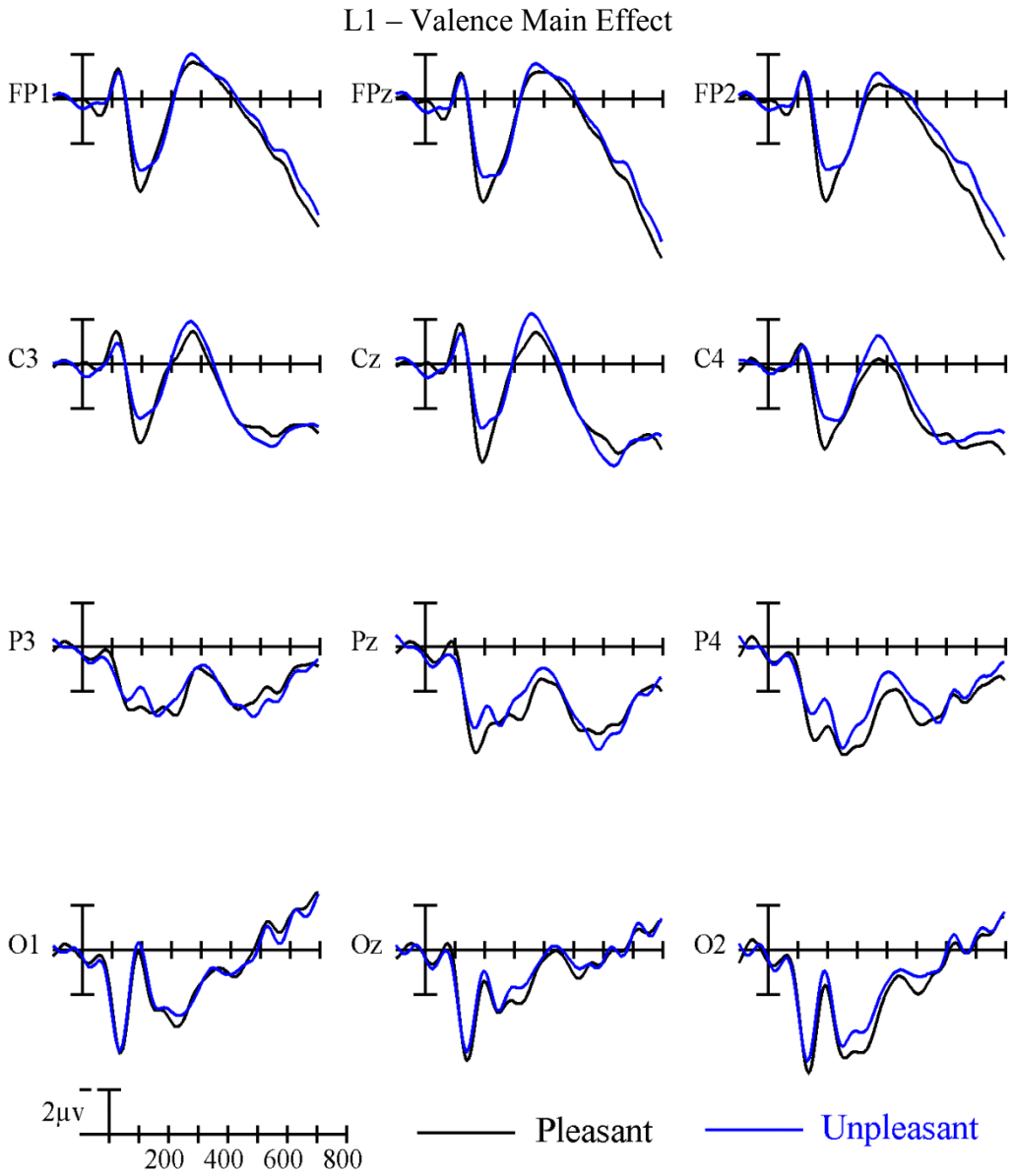

Figure 6a. ERPs to high valence (pleasant) and low valence (unpleasant) items in native English speakers.

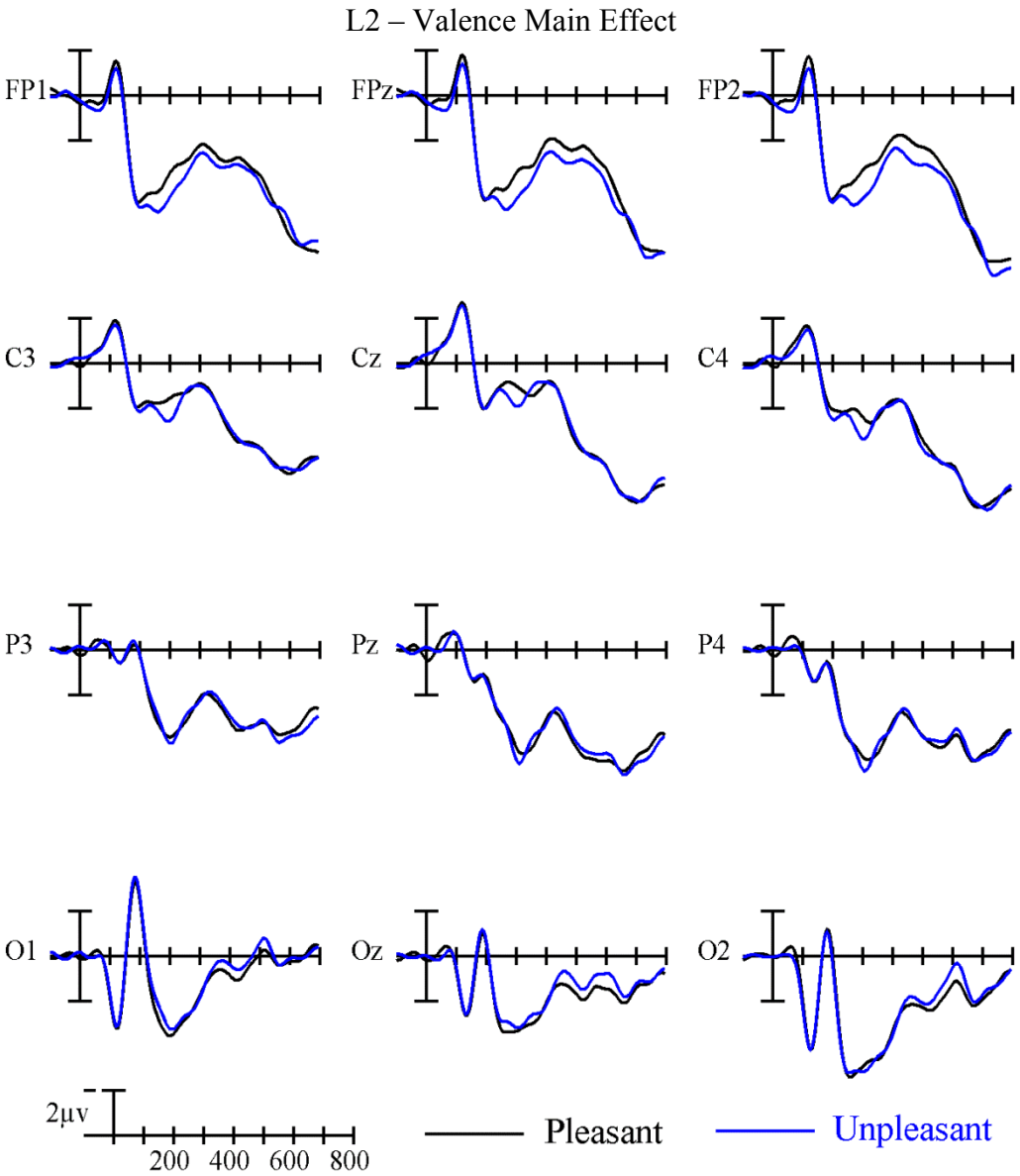

Figure 6b. ERPs to high valence (pleasant) and low valence (unpleasant) items in L2 English speakers.

Analysis of ERP Data

150-250 ms epoch.

No significant main effects of GROUP, VALENCE or AROUSAL were observed, though an interaction of VALENCE and AROUSAL was present ($F(1,30) = 9.1$, $p < 0.01$). The interaction of GROUP and VALENCE was significant ($F(1,30) = 4.5$, $p = 0.04$), as was the interaction of GROUP

x VALENCE x AROUSAL x POSTERIOR ($F(3,90) = 3.3$, $p = 0.045$). Follow-up analyses were conducted on the data from each group of participants.

Native speakers

The main effect of VALENCE was significant ($F(1,15) = 7.9$, $p = 0.01$). No main effect of AROUSAL or interaction between VALENCE and AROUSAL was observed.

L2 speakers

No main effect of VALENCE or AROUSAL was present. The interaction of VALENCE and AROUSAL was significant ($F(1,15) = 6.8$, $p = 0.02$).

300-500 ms epoch.

The main effect of GROUP approached significance ($F(1,30) = 3.8$, $p = 0.06$). No main effects of VALENCE or AROUSAL were present. The only significant interaction was that of GROUP x VALENCE x AROUSAL x LATERALITY ($F(2,60) = 7.6$, $p < 0.01$). Follow-up analyses were conducted on the data from each group of participants.

Native speakers

No significant main effects were observed. The interaction of VALENCE x AROUSAL x LATERALITY was present ($F(2,30) = 5.0$, $p = 0.03$).

L2 speakers

No significant main effects or interactions were observed.

500-700 ms epoch.

A late positive component (LPC) was evident around 500-700 ms. An ANOVA revealed a trend toward a main effect of GROUP ($F(1,30) = 3.5$, $p = 0.07$). Across the two groups, the main effect of AROUSAL was significant ($F(1,30) = 4.2$, $p = 0.048$) and an interaction of VALENCE and AROUSAL was present ($F(1,30) = 4.6$, $p = 0.04$). An interaction of GROUP x AROUSAL was not observed ($F(1,30) = 1.4$, $p > 0.10$). The only interaction that varied significantly by group was that of GROUP x VALENCE x AROUSAL x LATERALITY ($F(2,60) = 7.7$, $p < 0.01$). Follow-up analyses were conducted on each group of participants.

Native speakers

A main effect of AROUSAL was present across the region of analysis, with high arousal words eliciting a greater positivity than low arousal words ($F(1,15) = 5.8$, $p = 0.03$). There was no main effect of VALENCE, but the interaction between VALENCE and LATERALITY reached significance ($F(2,30) = 4.3$, $p = 0.03$). No interaction between VALENCE and AROUSAL was observed, though the interactions of VALENCE x AROUSAL x POSTERIOR ($F(3,45) = 2.9$, $p = 0.06$) and VALENCE x AROUSAL x LATERALITY ($F(2,30) = 3.6$, $p = 0.06$) both approached significance.

L2 speakers

No main effects of VALENCE or AROUSAL were observed. The interaction of VALENCE x AROUSAL did not reach significance across the analyzed region ($F(1,15) = 3.5$, $p = 0.08$), but an interaction of VALENCE x AROUSAL x LATERALITY was present ($F(2,30) = 5.5$, $p = 0.02$).

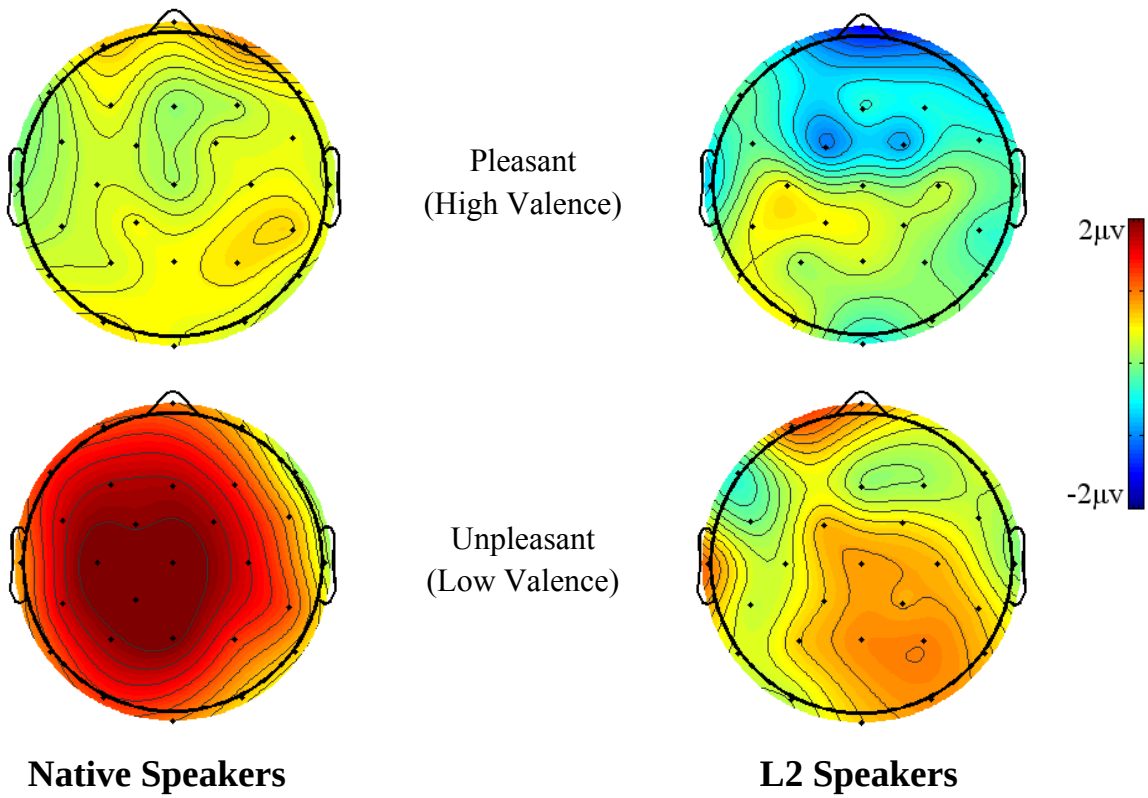

Figure 7. Voltage maps of low arousal subtracted from high arousal words within each level of valence from 500-700 ms.

Behavioral Data

L1 participants detected on average 96.8% (*SD* = 3.1) of probe items with a reaction time of 654 ms (*SD* = 54.4). L2 participants detected significantly fewer probes (mean = 91.3%, *SD* = 11.3; $t(30) = -1.9$, $p = 0.03$) but did not do so significantly slower than the L1 group (mean = 695 ms, *SD* = 101.0; $t(30) = -1.4$, $p = 0.08$). L1 participants produced false alarms on an average of 0.56 items (*SD* = 0.63) and L2 on an average of 0.81 items (*SD* = 0.98), a difference that was not statistically significant.

Discussion

Experiment 1 predicted that arousal would affect the LPC in native speakers of English but not in L2 speakers. This hypothesis was confirmed; the difference between high and low arousal words was significant in L1 speakers, while this effect was not observed in L2 speakers. The observed group differences were confirmed by an analysis revealing a significant interaction of GROUP x VALENCE x AROUSAL x LATERALITY, indicating that L2 speakers of a language do not process emotional attributes of a words in the same way as native speakers. The absence of a statistically significant interaction between GROUP and AROUSAL was unexpected given the differences between the observed arousal effects in each group and may be due to a lack of power due to the small sample size or to variance in L2 speakers as a result of varying language histories. The L2 English group in Experiment 1 included both early and late bilinguals; previous studies have found differences in these populations' responses to affective language stimuli (Harris et al., 2004; Caldwell-Harris & Ayçiçeği-Dinn, 2009).

The finding of an interaction of valence and arousal in both groups of participants was unexpected based on the results of Delaney-Busch et al. (2011), who found equal effects of arousal on the LPC across multiple levels of valence. In the present study, arousal effects were diminished in high valence (pleasant) words compared to low valence (unpleasant) words in monolinguals. Though much less robust, the same pattern was observed in L2 speakers with a different laterality of effects. This result is consistent with those of the only previous study that has examined the influence of emotion on the LPC using a semantic categorization task (Fischler & Bradley, 2006). Though the stimuli in the present study were taken from those used by Delaney-Busch et al., they represent only a subset of their items. Our stimuli were therefore not as rigorously matched for levels of concreteness and valence across the four crossed emotional

conditions. However, given the magnitude of our observed effects, it seems unlikely that these differences are solely responsible for the present results.

Experiment 2

Experiment 1 found that L2 speakers of a language did not exhibit the arousal effect on the LPC seen in native speakers. Experiment 2 aimed to extend this finding to the domain of taboo words. As the heterogeneity of the bilingual group's language experiences in Experiment 1 may have limited the interpretability of results, the L2 English group in Experiment 2 consisted of only late bilinguals.

Methods

Participants

32 Tufts University students (20 female, mean age = 20.4, *SD* = 2.2) were recruited and paid for their participation. All participants were right handed, had normal or corrected-to-normal vision and reported no history of traumatic head injury. Sixteen participants had learned English as their L2 after the age of 6, beginning at a mean age of 8.2 years (range = 6-13 years, *SD* = 2.48). Participants in this group spoke twelve different languages (Japanese, French, Mandarin Chinese, Bengali, Vietnamese, Portuguese, Turkish, Malayalam, Nepali, Russian, Spanish and Hindi) as their L1. The remaining 16 participants were native English speakers who had not been exposed to any other language before the age of 6 and had not had any language learning experience in the past two years.

Participants rated their abilities to read, speak and comprehend their L1 and L2 as well as how frequently they read and communicated in each language on a 7-point Likert scale (1 =

unable to 7 = *expert*). Native English speakers' average level of English ability was 6.9 (SD = 0.36). Participants who learned English as a L2 rated their ability significantly lower at 6.1 (SD = 1.02) ($t(126) = 5.75$, $p < 0.001$). Both groups reported that English was the language they used most frequently; on a 7-point Likert scale (1 = *rarely*; 7 = *very frequently*), L1 English speakers rated their frequency of use as 6.8 (SD = 0.54) and L2 English speakers reported an English use frequency of 6.6 (SD = 0.84) ($t(62) = 1.24$, $p > 0.20$).

Stimuli

The stimuli used in this study were 287 English words. Each item used in the present experiment was between three and thirteen letters, and the mean length of all items was 6.7 letters (SD = 1.86). The average frequency of the words as occurring in the Hyperspace Analogue to Language (HAL) frequency norms was 99.6 occurrences per million (SD = 257.59). 240 of these items were the same as the critical items in Experiment 1, with the exception of five items that were changed for better control of properties (see Table 2 for characteristics of these items; see Appendix A for complete list of critical items). High arousal items significantly differed from low arousal items on the measure of arousal ($t(158) = 27.83$, $p < 0.001$) and did not significantly differ in valence, concreteness, length or frequency (all ps > 0.1). High valence items significantly differed from low valence items on the measure of valence ($t(158) = 57.53$, $p < 0.001$) and did not significantly differ in arousal, concreteness, length or frequency (all ps > 0.2). Forty neutral items with medium ratings of both valence and arousal were included to provide a means of comparing the emotional items to non-emotional items. An additional 40 items with relatively neutral valence and arousal ratings served as fillers to make the experimental manipulation less evident. The taboo condition contained 47 three to eight letter taboo words identified by the experimenters. Items in the taboo condition could not be matched

to the other conditions in terms of length due to the limited English taboo lexicon, but taboo words were comparable to neutral words in written frequency ($t(85) = 0.33$, $p = 0.74$).

	Valence	Arousal	Concreteness	Length	Frequency
VA	5.71 *(0.32)*	5.06 (*0.38)*	3.54 *(0.93)*	7.08 (*1.76)*	76.53 (*132.53)*
Va	5.53 *(0.41)*	3.16 *(0.45)*	3.80 *(1.21)*	6.83 *(2.12)*	190.08 *(496.49)*
vA	1.92 *(0.39)*	4.97 *(0.45)*	3.73 *(1.02)*	7.05 *(1.81)*	53.75 *(99.59)*
va	2.28 *(0.32)*	3.17 *(0.40)*	4.02 *(1.03)*	6.35 *(2.09)*	106.43 *(237.98)*
Neutral	3.62 *(0.74)*	4.10 *(0.09)*	3.85 *(0.77)*	7.20 *(1.54)*	95.04 *(178.45)*
Taboo				5.43 *(1.53)*	82.73 *(171.73)*

Table 2. Mean valence, arousal, concreteness, length, and frequency (occurrences per million) of words used in the experiment.

Forty-five probe items of four to eleven letters in the semantic category of "animal names" (such as *cat* or *dog*) were included in the presented lists (14% of all items). Each participant saw every experimental and probe item in one of six randomly sorted orders.

Procedure

The procedure for this experiment was the same as in Experiment 1. After electrode placement, instructions for the experimental task were given and a short practice list was presented to familiarize participants with the task. The experimental section typically took 20 minutes to complete and contained three pauses of a length determined by the participant. After the ERP experiment was completed, participants were asked to rate the tabooness of each of the 287 experimental items on a 7-point Likert scale (1 = *not taboo*; 7 = *highly taboo*).

EEG recording

The EEG recording procedure for this experiment was the same as in Experiment 1.

Data analysis

Averaged ERPs were formed off-line from trials free of ocular and muscular artifact (13% of all trials were excluded) and were lowpass filtered at 15 Hz. The same subset of 12 of the 29 scalp electrodes as in Experiment 1 was chosen for analysis. Mean amplitude was measured in the epochs of 150-250, 300-500, and 500-700 ms. Average waveforms were created for two levels of GROUP, two levels of TABOO (Taboo and Neutral), three levels of LATERALITY, and four levels of POSTERIOR. Only trials containing items that each participant personally rated as having a high level of tabooness were averaged in the Taboo condition; the minimum rating for a word to be included varied according to the distribution of each subject's ratings. In a separate analysis, average waveforms were also created for two levels of GROUP, two levels of VALENCE (High and Low), two levels of AROUSAL (High and Low), three levels of LATERALITY, and four levels of POSTERIOR, as in Experiment 1. Repeated analyses of variance (ANOVAs) were performed on the data in each epoch. The Geisser and Greenhouse (1959) correction was applied to all repeated measures with more than one degree of freedom in the numerator.

Results

Visual Inspection of ERPs

Grand-average ERP waveforms elicited by taboo and neutral words for each participant group are shown in Figure 8a and 8b. Voltage maps created by subtracting the ERPs for neutral words from the ERPs for taboo words for each language group are shown in Figure 9. Grand-

average ERP waveforms elicited by high and low arousal words for each participant group are shown in Figures 10a and 10b. Figure 11 shows voltage maps created by subtracting the ERPs for low arousal items from the ERPs for high arousal items for each language group. ERP waveforms elicited by high and low valence words for each participant group are shown in Figures 12a and 12b. As can be seen in Figures 8a and 8b, the first visible component was a small negativity between 90 and 150 ms after stimulus onset (N1), followed by a positivity peaking around 200 ms (P2). At frontal sites the negativity between 300 and 500 ms (N400) peaks approximately 50-100 ms later for taboo than for neutral words in both groups. The positivity between 500 and 700 ms (LPC) is more pronounced for taboo than for neutral words, with differences occurring across sites in the L1 group and at centro-posterior sites in the L2 group.

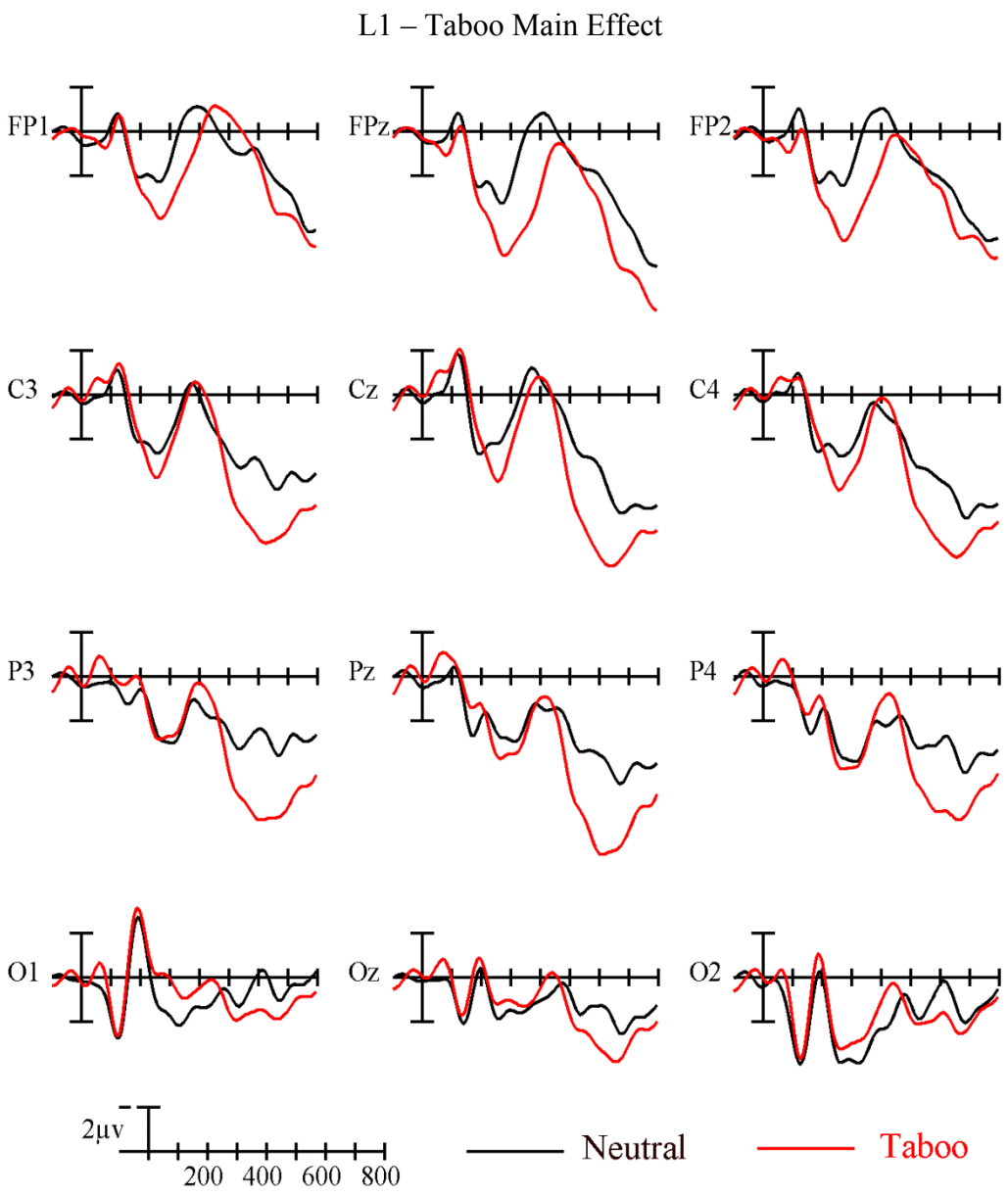

Figure 8a. ERPs to taboo and neutral items in native English speakers.

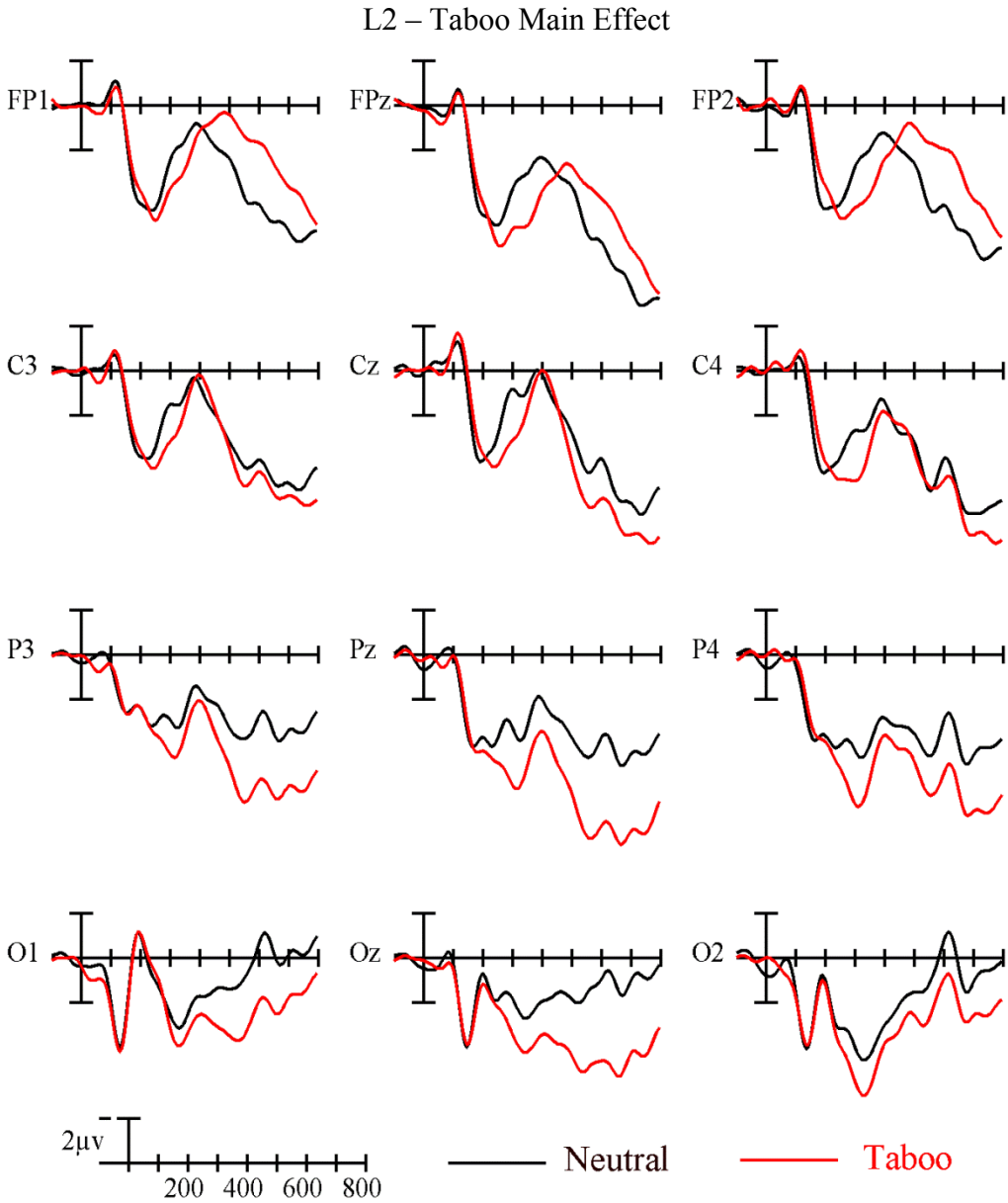

Figure 8b. ERPs to taboo and neutral items in L2 English speakers.

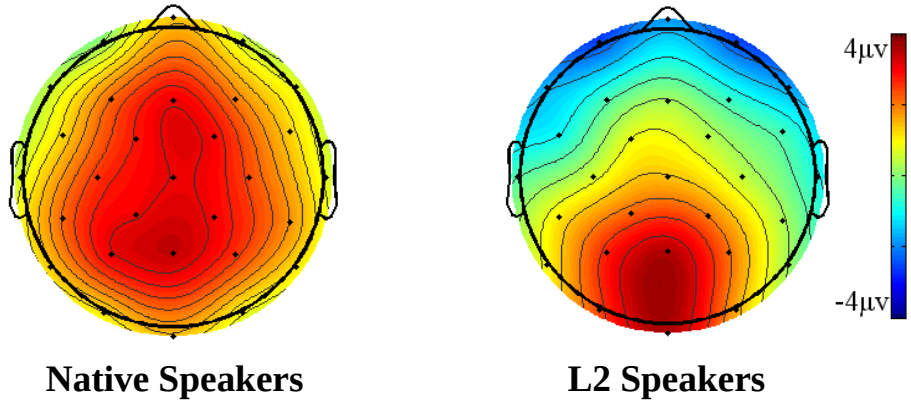

Figure 9. Voltage maps of neutral subtracted from taboo words from 500-700 ms.

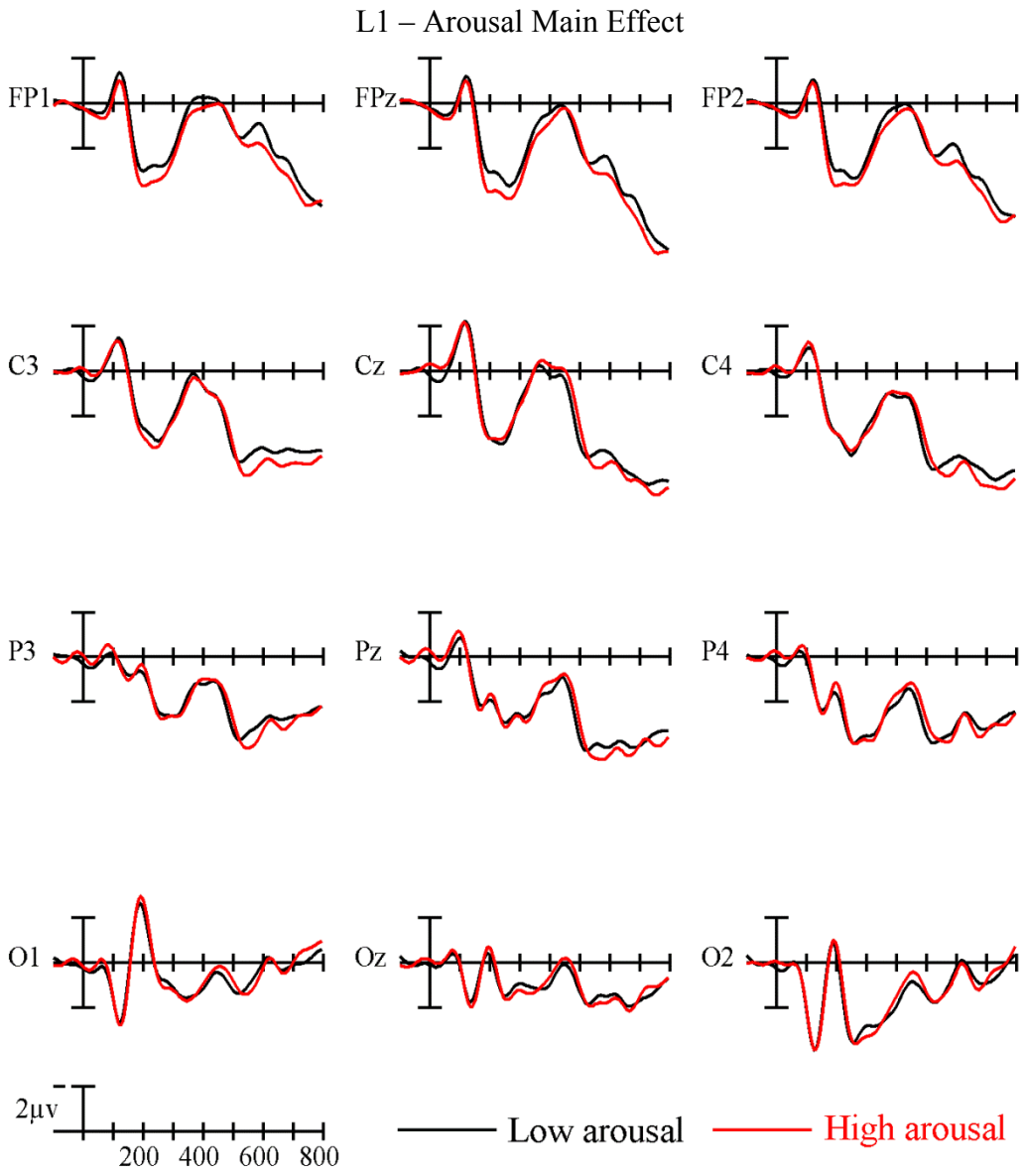

Figure 10a. ERPs to low and high arousal items in native English speakers.

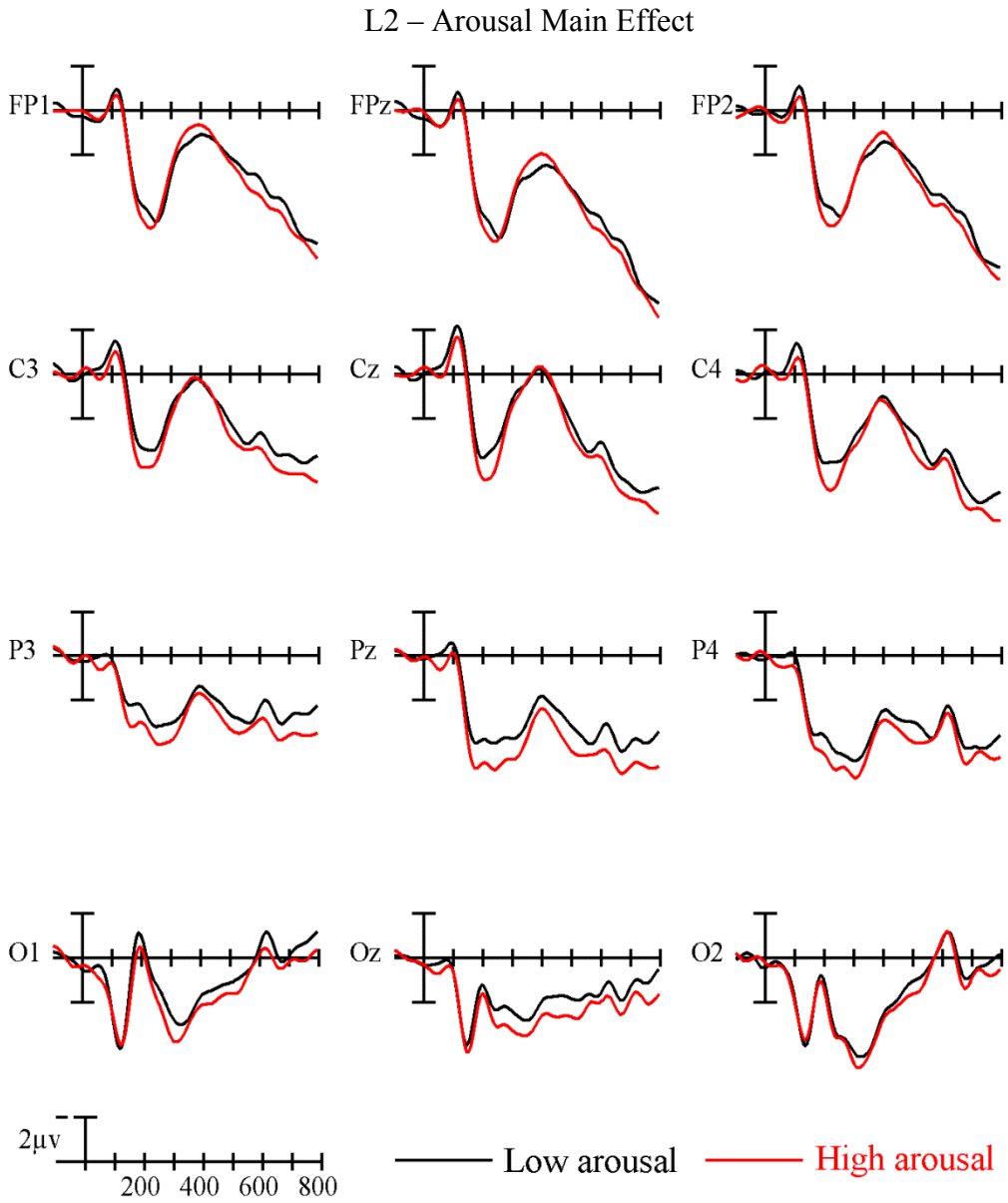

Figure 10b. ERPs to low and high arousal items in L2 English speakers.

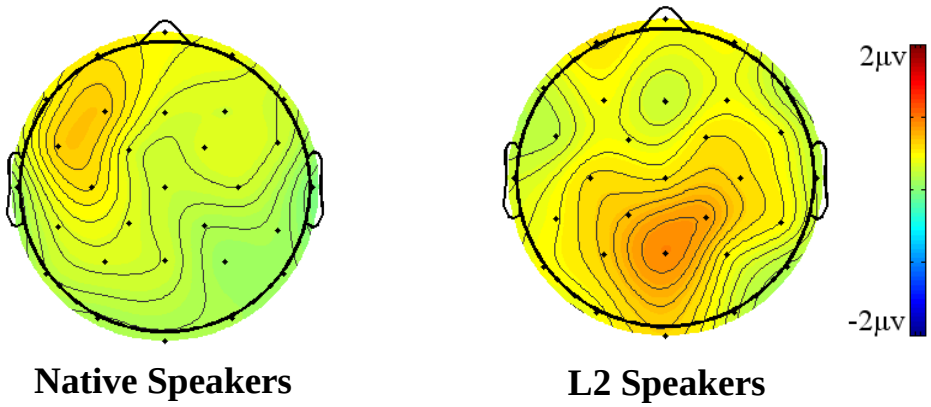

Figure 11. Voltage maps of low arousal subtracted from high arousal items from 500-700 ms.

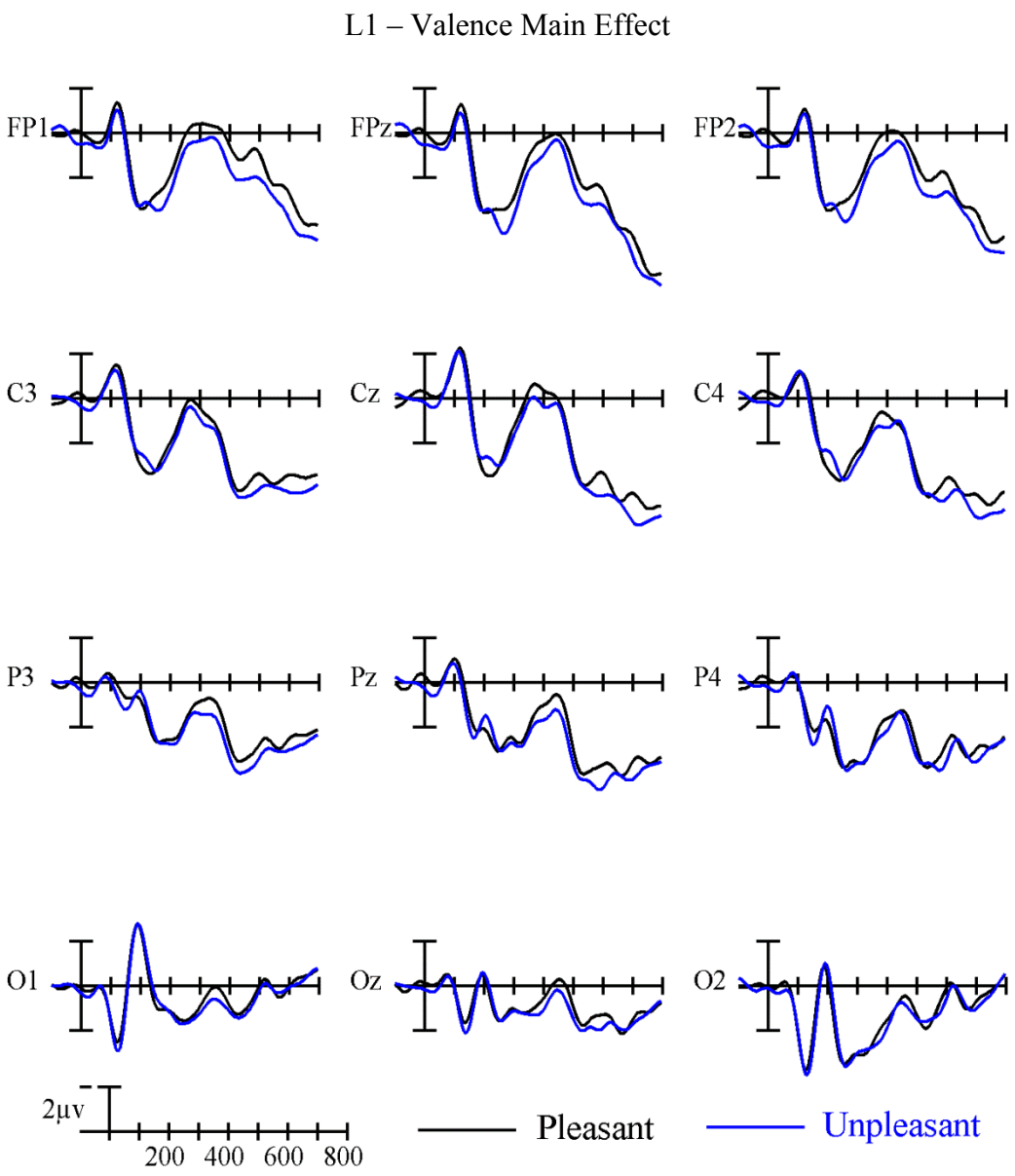

Figure 12a. ERPs to high valence (pleasant) and low valence (unpleasant) items in native English speakers.

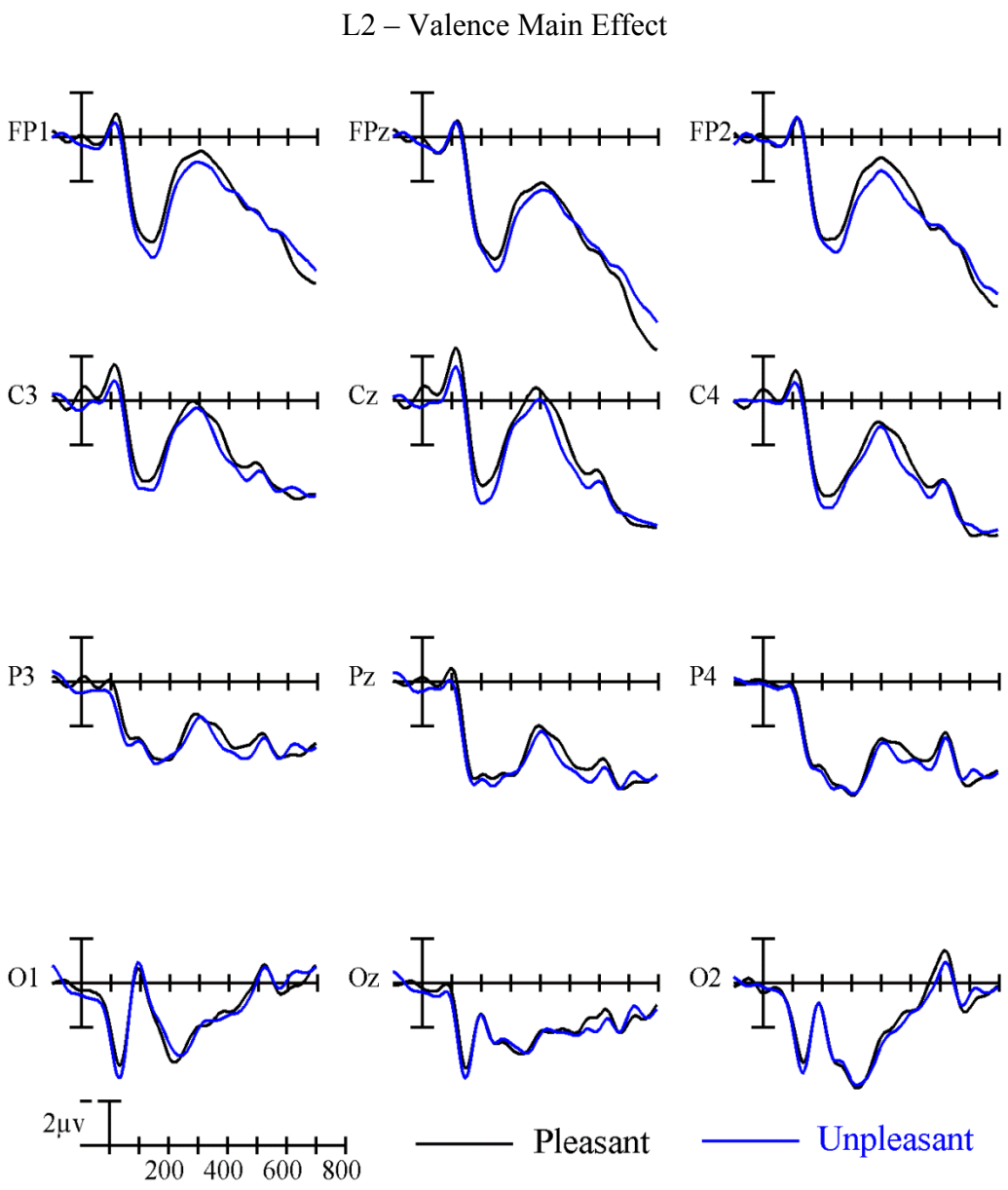

Figure 12b. ERPs to high valence (pleasant) and low valence (unpleasant) items in L2 English speakers.

Analysis of ERP Data

150-250 ms epoch.

In the TABOO analysis, there was a main effect of GROUP ($F(1,30) = 7.2$, $p = 0.01$). No main effect of TABOO was observed, but the interaction of GROUP x TABOO x POSTERIOR was significant ($F(3,90) = 4.0$, $p = 0.03$).

In the comparison of VALENCE and AROUSAL, a main effect of GROUP was observed ($F(1,30) = 7.5$, $p = 0.01$). VALENCE and AROUSAL did not produce significant main effects or interactions other than the interaction of GROUP x AROUSAL x POSTERIOR, which approached significance ($F(3,90) = 3.3$, $p = 0.05$). Given that group differences were observed, follow-up analyses were conducted on the data from each group of participants.

Native speakers

TABOO interacted significantly with POSTERIOR ($F(3,45) = 7.0$, $p < 0.01$). No main effects of VALENCE or AROUSAL were present, though an interaction of AROUSAL and POSTERIOR was observed ($F(3,45) = 4.0$, $p = 0.04$).

L2 speakers

No effects of TABOO were observed. There was no main effect of VALENCE, but AROUSAL was significant ($F(1,15) = 5.9$, $p = 0.03$). No significant interactions were present.

300-500 ms epoch.

Across groups, there was a main effect of TABOO ($F(1,30) = 8.6$, $p = 0.01$). The main effect of GROUP did not reach significance ($F(1,30) = 3.2$, $p = 0.08$), but the interaction of GROUP x TABOO x POSTERIOR was significant ($F(3,90) = 5.1$, $p = 0.02$).

In the comparison of VALENCE and AROUSAL, a main effect of VALENCE was observed ($F(1,30) = 4.2$, $p = 0.048$). A main effect of AROUSAL was not present, but AROUSAL interacted

with GROUP and POSTERIOR ($F(3,90) = 3.9$, $p = 0.03$). Follow-up analyses were conducted on the data from each group of participants.

Native speakers

The main effect of TABOO was not significant, but TABOO interacted significantly with both POSTERIOR ($F(3,45) = 5.0$, $p = 0.03$) and LATERALITY ($F(2,30) = 5.9$, $p = 0.02$). No significant effects of VALENCE or AROUSAL were observed.

L2 speakers

The main effect of TABOO was significant ($F(1,15) = 6.3$, $p = 0.02$). No significant effects of VALENCE or AROUSAL were observed.

500-700 ms epoch.

A late positive component (LPC) was evident beginning around 500 ms. The LPC was assessed using a 500-700 ms time window, consistent with Experiment 1. Across groups, the main effect of TABOO was significant ($F(1,30) = 11.5$, $p < 0.01$). The main effect of GROUP was not significant, and GROUP x TABOO did not significantly interact (ps > 0.2). However, an interaction of GROUP x TABOO x POSTERIOR was observed ($F(3,90) = 7.8$, $p < 0.01$).

No main effects of VALENCE or AROUSAL were present, but the interaction of GROUP x VALENCE x AROUSAL was significant ($F(1,30) = 4.5$, $p = 0.04$). Follow-up analyses were conducted on the data from each group of participants.

Native speakers

A significant main effect of TABOO was observed across the region of analysis ($F(1,15) = 12.8$, $p < 0.01$). Additionally, the interactions between TABOO x LATERALITY ($F(2,30) = 4.2$, $p = 0.04$) and TABOO x POSTERIOR ($F(3,45) = 7.4$, $p < 0.01$) were significant.

No main effect of AROUSAL or interactions of AROUSAL with LATERALITY or POSTERIOR were observed (all *p*s > 0.40). Additionally, no main effect of VALENCE or interactions with LATERALITY or POSTERIOR were present (all *p*s > 0.10). The interaction of VALENCE and AROUSAL approached significance ($F(1,15) = 3.9$, $p = 0.07$) and did not interact with any dimensions of location.

L2 speakers

The main effect of TABOO was not significant across the region of analysis ($F(1,15) = 2.2$, $p = 0.16$); however, the interaction of TABOO x LATERALITY ($F(2,30) = 9.3$, $p < 0.01$) and that of TABOO x POSTERIOR ($F(3,45)= 18.1$, $p < 0.001$) were both highly significant.

No main effect of AROUSAL or interactions of AROUSAL with LATERALITY or POSTERIOR were obtained (all *p*s > 0.1). There was no main effect of VALENCE or interactions with LATERALITY or POSTERIOR (all *p*s > 0.5). An interaction of VALENCE x AROUSAL was not observed.

Behavioral Data

L1 participants detected on average 96.1% of probes ($SD = 3.7$) with a mean reaction time of 708 ms ($SD = 77.9$). L2 participants detected significantly fewer probes (mean = 81.4%, $SD = 14.9$; $t(30) = -3.8$, $p < 0.01$) and took significantly longer to do so (mean = 760, $SD = 79.4$;

$t(30) = -1.9$, $p = 0.04$). L1 participants produced false alarms on an average of 1.5 items (*SD* = 2.1), less than L2 participants who had an average of 3.3 items (*SD* = 3.2) ($t(30) = 1.9$, $p = 0.03$).

L1 participants rated the words in the Taboo condition as having an average tabooness of 5.0 (*SD* = 1.01). L2 participants rated the words in this condition as being significantly less taboo, with an average tabooness rating of 4.1 (*SD* = 1.16) ($t(46) = -8.76$, $p < 0.001$). Only two of the 47 items, both relatively "mild" words, were rated as having a higher tabooness by the L2 speakers.

Discussion

In Experiment 2, taboo words elicited a large LPC. This pattern resembles the effect found to highly arousing words in Experiment 1, suggesting that reactions to taboo words are a function of their high levels of arousal. As predicted by the results of a significant arousal effect in L1 but not L2 speakers in Experiment 1, the taboo effect was only significant across the region of analysis in L1 speakers, supporting the hypothesis that late bilinguals experience a diminished impact of taboo words in their L2. This conclusion is bolstered by behavioral data, which found that L2 speakers rate the "tabooness" of taboo words lower than monolingual native speakers. However, the effect of taboo words in a L2 was not completely negligible; the interactions between taboo and dimensions of locality were significant in L2 speakers, indicating that these words showed a smaller distribution of effects. This suggests that taboo words are not powerless even to late learners of a language. Surprisingly, no effects of arousal on nontaboo items were observed in Experiment 2.

Comparison of Experiments

Experiment 1 revealed a robust effect of arousal on the LPC in native English speakers. Though the nontaboo emotional stimuli used were largely identical, this effect was absent in Experiment 2, which contained taboo words in addition to the nontaboo emotional items. The striking disparity between these results (see Table 3 and Table 4 for a summary of significant results in each experiment) prompted a direct comparison of the L1 groups from the two experiments.

	Epoch:	150-250	300-500	500-700
Overall	Exp 1	VA, GV, GVA*	G, GVA*	G, A, VA, GVA*
	Exp 2	G	V, GA*	GVA
Native Speakers	Exp 1	V	VA*	A, V*, VA*
	Exp 2	A*	!	!
L2 Speakers	Exp 1	VA	!	VA*
	Exp 2	A	!	!

Table 3. Significant results ($p < 0.05$) in each experiment for GROUP (G), VALENCE (V), and AROUSAL (A). * indicates an interaction with LATERALITY or POSTERIOR.

Epoch:	150-250	300-500	500-700
Overall	G, GT*	T, GT*	T, GT*
Native Speakers	T*	T*	T, T*
L2 Speakers	!	T	T*

Table 4. Significant results ($p < 0.05$) in Experiment 2 for TABOO (T) and GROUP (G). * indicates an interaction with LATERALITY or POSTERIOR.

Data analysis

The same set of 12 electrodes as in Experiments 1 and 2 were analyzed. Mean amplitude was measured in the LPC epoch of 500-700 ms. Average waveforms were created for two levels of EXPERIMENT (Experiment 1 and Experiment 2), two levels of VALENCE (High and Low), two

levels of AROUSAL (High and Low), three levels of LATERALITY, and four levels of POSTERIOR. Repeated analyses of variance (ANOVAs) were performed on the data in each epoch. The Geisser and Greenhouse (1959) correction was applied to all repeated measures with more than one degree of freedom in the numerator.

Results

Analysis of ERP Data: 500-700 ms epoch.

No main effect of EXPERIMENT was observed ($F(1,30) = 1.0$, $p > 0.32$). The interaction of EXPERIMENT and AROUSAL was not significant, but an interaction of EXPERIMENT x VALENCE x AROUSAL was found ($F(1,30) = 4.2$, $p < 0.05$).

Behavioral Data

There were no significant differences between the two L1 groups' numbers of probes detected or false alarm rates (all *p*s > 0.10). However, the presence of taboo words affected L1 participants' reaction times; they took on average 654 ms (*SD* = 54.4) to react to probes in Experiment 1, but responded on average over 50 ms slower in Experiment 2 (mean = 708 ms, *SD* = 77.9; $t(30) = -2.3$, $p = 0.03$). The same effect was not observed in L2 participants, who reacted in an average of 695 ms (*SD* = 101.0) in Experiment 1 and 760 ms (*SD* = 79.4) in Experiment 2 ($t(30) = -2.0$, $p = 0.05$).

Discussion

Though the interaction of EXPERIMENT and AROUSAL was not significant, an interaction of EXPERIMENT x VALENCE x AROUSAL was observed. This indicates that the presence of taboo words in the experimental paradigm was sufficient to modify monolingual native English

speakers' processing of nontaboo emotional words. This finding will be discussed further in the following General Discussion.

General Discussion

The present study aimed to examine how emotional and taboo words are processed differently in L2 than in L1. Previous investigations into the subject have varied widely in their findings, perhaps due to inadequate sensitivity of the measures used or to great differences in the populations studied. In Experiment 1, native and L2 speakers of English were presented with words crossed in levels of valence and arousal. Based on previous research showing that the late positive component (LPC) is triggered by arousal in monolinguals, it was predicted that L2 English speakers would show attenuated effects of arousal on the LPC compared to native speakers. Taboo words were added to the paradigm in Experiment 2. We anticipated that taboo word effects would follow a similar pattern to those of highly arousing words, eliciting an LPC that would be more positive and more widely distributed in native speakers than in L2 speakers.

In Experiment 1, a more positive LPC was elicited by high arousal than by low arousal words in native English speakers, an effect that was greatest at centro-posterior sites. In contrast, this pattern of effects was not observed in L2 speakers, indicating that arousal affected processing in native speakers but not in L2 speakers. There was also an unexpected interaction of valence and arousal, inconsistent with the findings of Delaney-Busch et al. (2011), who found only a main effect of arousal on the LPC. It is possible that this disparity can be attributed to the less precise matching of stimuli across categories in the present experiment. However, as will be discussed below, changing the context in which words appear could have caused them to be experienced as subjectively higher or lower in arousal; the fact that our words appeared in a list

with a different composition than that used by Delaney-Busch et al. could be responsible for the difference.

In Experiment 2, we observed that the pattern of results to taboo words was similar to the arousal effects observed in experiment 1; taboo words elicited a large LPC. Moreover, while in native English speakers the effects were robust across the scalp, the LPC elicited by taboo words was restricted in spatial distribution for L2 speakers. However, this effect was not entirely negligible, indicating that taboo words do still have an impact in L2. These results, the first to investigate the emotional characteristics of taboo words with ERPs, demonstrate that these words are essentially processed like highly arousing items. Unexpectedly, no effects of arousal were found to highly arousing nontaboo words in Experiment 2.

We did not anticipate observing any effects in earlier epochs than the LPC (150-250 and 300-500 ms). There have been reports in the literature of an early posterior negativity (EPN) evoked around 200-300 ms after presentation of an emotional word, similar to that seen in response to emotionally arousing pictures (Kissler et al., 2006; Herbert et al., 2008). Though there is not a consensus on precisely when this effect occurs or even if it is likely that much semantic processing occurs during this early stage, the EPN has generally been found to be responsive to both pleasant and unpleasant compared to neutral words, suggesting that it may be triggered by arousal (Kissler, Herbert, Peyk, & Junghofer, 2007; Kissler et al., 2009). However, the effects in the present study do not reliably occur in one time window that could be said to represent this component. Due to the restrictions of selecting words with specific levels of valence and arousal, it is possible that our items could have varied on some low-level characteristic that influenced early effects. As it is unclear what these effects might signify in the

context of our focus on L1 vs. L2 word processing, the remaining portion of this discussion will be limited to results observed in our epoch of interest, the LPC.

The most surprising finding of the present study was that the addition of taboo words in Experiment 2 to a nearly identical list from Experiment 1 changed the effect of arousal on nontaboo items. Unlike in Experiment 1, no arousal effects were observed to nontaboo items in Experiment 2. While this difference was unexpected, it is not unprecedented; Harris et al. (2003) reported a similar observation that nontaboo emotion effects in skin conductance responses became negligible when taboo words were added to an experiment. However, to the extent of our knowledge the present study is the first to observe this phenomenon using ERPs. Arousal is often described as being processed in a highly automatic manner, implying that arousal levels are an inherent, immutable property of words (Kissler et al., 2006, 2007). Given that arousal levels of taboo words are much higher than those of nontaboo emotional words, this study demonstrates that arousal may be locally relative and dependent on context. As it has been established that the LPC is diminished when participants do not attend to the semantic content of words, it is possible that the meanings of emotional words were ignored when much more salient taboo words were present (Fischler & Bradley, 2006). Participants in both experiments were not provided any details of the manipulation during testing sessions, but participants in Experiment 2 were warned about the presence of potentially offensive words in the experiment during the initial stages of the recruitment process and when giving informed consent. Participants who remembered these warnings could have been attending to taboo words despite the lack of instructions to do so; it is also possible that taboo words are simply so unexpected in the academic environment of a psychology study that participants automatically treated them as a distinct category.

Reaction times (RTs) to probe items in the two experiments also support the idea that taboo words affected performance on nontaboo items. It has previously been established that RTs to a given trial can vary depending on the surrounding list context (Perea, Carreiras, & Grainger, 2004). In the present study native English speakers' responses to the same probes were significantly slower in Experiment 2 than in Experiment 1, suggesting that they may have been noticing taboo words as a parallel "task" that interfered with performance on the assigned semantic categorization task. Intriguingly, this effect was not evident in L2 speakers; despite the expectation that the L2 group in Experiment 2 would have longer RTs due to lower English proficiency and a lower percentage of correct responses than the L2 group in Experiment 1, there was not a significant difference in RTs to probe items between experiments. It seems that taboo words did not interfere with L2 speakers' performance to the same degree as native speakers', lending credence to our claim that taboo words have less impact in L2; if taboo words are less powerful, they do not have the same distracting force. Likewise, L2 speakers of English rated the tabooness of words in the taboo condition significantly lower than L1 speakers; their lower subjective judgments of tabooness also support our hypothesis of diminished impact.

These results demonstrate that emotional and taboo words are processed differently in a native language and in a second language, even when L2 was acquired at a relatively early age, substantiating the claim that words in L2 have less of an emotional impact than in L1. Our findings are not in line with those of previous studies that failed to demonstrate a difference in emotionality between L1 and L2. Differences in methodologies may account for this disparity. As highly arousing words are thought to trigger rapid, automatic reactions, some of the measures used by previous studies likely lacked the sensitivity to detect these initial differences. The memory and Stroop tasks employed by most of the previous research in this area, with dependent

measures of recall performance or reaction time, are not capable of measuring differences in the early stages of word processing. Colbeck and Bowers (2012), who did find a difference between L1 and L2, used a RSVP task that may be a better indicator of early automatic processing; in this paradigm stimuli are presented briefly, so deep processing cannot occur. Consistent evidence of language differences has also been found in studies measuring skin conductance responses, which index automatic emotional responses. The present study used a semantic categorization task, which requires participants to attend to the general meaning but not to any particular aspect of the presented items. While performing a semantic categorization task is undeniably very different from participating in everyday discourse, participants experience language in a way that is more naturalistic than in tasks like a Stroop task where they must try to ignore the meanings of words.

The only previous study to examine this issue using ERPs differed from the present investigation in several ways. Paramount is that Conrad et al. (2011) did not control for arousal in their stimuli, though this factor is thought to be the main basis of emotional word effects. Based on their results, they conclude that emotion is processed qualitatively the same in L1 and L2. However, they do seem to have found evidence of some disparity between languages: their group of native Spanish speakers had a significant LPC only to pleasant- and not to unpleasant-words compared to neutral in L2. The authors propose that this may represent a positivity bias resulting from the “positive life experience” of living in an L2 country; we suggest that, given that their pleasant and unpleasant items were not controlled for arousal, this may actually be indicative of an attenuated response to arousal in this L2 group (the less proficient of the two in their study). It is specified that their stimuli were controlled for valence across languages, but the process of doing so may have actually reduced the similarity of arousal levels between

categories, as it is unlikely that valence and arousal would correlate exactly. Additionally, Conrad et al. (2011) compared the same group of participants reading translations of words in two languages, not two groups of participants both reading identical items as in the present study. While a within-subjects design eliminates the problem of individual differences between groups, it requires a comparison to be made across two different sets of stimuli. The authors do not discuss the differences between their groups' language histories at length, but it is stated that native German speakers learned their L2 at an earlier age and for a longer period of time while having spent less time in an L2-speaking country. In contrast, native Spanish speakers began learning their L2 at a later age and spent less time doing so, most of which was spent in an L2-speaking country. It seems that these groups had very different L2 learning environments and experiences.

At present, it cannot be stated with precision which elements of language learning and experience influence the emotionality of words in L2. A variety of factors have been proposed to be the crucial determinant of emotional impact, including L2 learning experience, age of L2 acquisition, L2 proficiency, and length of residency in an L2-speaking location. The present study provides some evidence for the influence of the environment in which L2 was learned, particularly in Experiment 2 where the L2 group consisted of entirely late bilinguals. Most late L2 learners grew up speaking L1 at home and learned English (L2) only in the classroom, although taboo words (the effects of which were diminished in L2) were presumably learned in less formal situations. However, we cannot substantiate any hypotheses regarding the importance of the age at which a language was acquired on its emotionality; the L2 population in Experiment 1 contained both early and late bilinguals, though it can be noted that differential effects from native speakers were still observed with this mixed group. Proficiency also cannot be eliminated

as a significant factor by our results. Though all L2 participants were highly proficient in English and current students in an American university, their self-reported English proficiency was significantly lower than that of native speakers in both experiments. While it is impossible to determine if differences in cultural attitudes could have influenced the responses of participants in the L2 group, the present study minimized the effects of this factor by the representation of a wide variety of backgrounds in our L2 population, unlike any previous research.

As in previous studies, language learning variables such as proficiency and age of acquisition were generally correlated in our participants. It is impossible to determine which factors have the greatest impact on L2 emotionality without performing controlled experiments. Studies could examine L2 emotion effects in, for example, groups of bilinguals who began learning L2 at different ages but have achieved the same level of proficiency, or groups who began learning L2 at the same age but in different environments. ERPs may be uniquely suitable as a technique to investigate these factors, as this method provides the degree of sensitivity to the time-course of visual word processing that is necessary in order to observe subtle differences.

It seems that arousal can be represented automatically through the medium of a single word, which only symbolically represents the intrinsic emotions from which arousal ultimately originates. The possibility that these effects might be represented early in the time-course of word processing, before even the commonly accepted threshold of semantic integration, is intriguing. Further studies are required to determine what our early effects may signify. If the difference between a first and a second language might be fundamental enough to alter these kinds of basic mechanisms, the aforementioned studies would be all the more enlightening to perform in order to determine which factors govern the emotionality of L2.

References

Anooshian, L. J., & Hertel, P. T. (1994). Emotionality in free recall: Language specificity in bilingual memory. *Cognition and Emotion, 8,* 503-514.

Arnell, K. M., Killman, K. V., & Fijavz, D. (2007). Blinded by emotion: Target misses follow attention capture by arousing distractors in RSVP. *Emotion, 7,* 465–477.

Ayçiçeği, A. & Harris, C. (2004). Bilinguals' recall and recognition of emotion words. *Cognition & Emotion, 18(7),* 977-987.

Bond, M.H. & Lai, T. (1986). Embarrassment and code-switching into a second language. *Journal of Social Psychology, 126,* 179-186.

Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Stimuli, instruction manual, and affective ratings*. (Tech. Rep. C-1). University of Florida, Gainesville, Center for Research in Psychophysiology.

Caldwell-Harris, C.L. & Ayçiçeği-Dinn, A. (2009). Emotion and lying in a non-native language. *International Journal of Psychophysiology, 71*, 193-207.

Colbeck, K. & Bowers, J. (2012). Blinded by Taboo Words in L1 but Not L2. *Emotion*. Advance online publication.

Conrad, M., Recio, G., & Jacobs, A.M. (2011). The time course of emotion effects in first and second language processing: a cross cultural ERP study with German–Spanish bilinguals. *Front. Psychology, 2*, 351.

Delaney-Busch, N., Haime, V., Wilkie, G., & Kuperberg, G.R. (2011). Vivid: A fully crossed ERP investigation of valence, salience, concreteness, and frequency. Annual Meeting of the Cognitive Neuroscience Society, 2011.

Dewaele, J.-M. (2004). The emotional force of swearwords and taboo words in the speech of multilinguals. *Journal of Multilingual and Multicultural Development, 25,* 204–222.

Dewaele, J.-M., & Pavlenko, A. (2002). Emotional vocabulary in interlanguage. *Language Learning, 52,* 263-322.

Eilola, T. M., Havelka, J., & Sharma, D. (2007). Emotional activation in the first and second language. *Cognition and Emotion, 21,* 1064–1076.

Ferré, P., Garcia, T., Fraga, I., Sanchez-Casas, R., & Molero, M. (2010). Memory for emotional words in bilinguals: Do words have the same emotional intensity in the first and in the second language? *Cognition and Emotion, 24,* 760–785.

Fischler, I., & Bradley, M. (2006). Event-related potential studies of language and emotion: words, phrases, and task effects. *Progress in Brain Research, 156*, 185–203.

Geisser, S., & Greenhouse, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24,* 95-112.

Gonzalez-Reigosa, F. (1976) The anxiety arousing effect of taboo words in bilinguals. In C. Spielberger and R. Diaz-Guerrero (eds) Cross-Cultural Anxiety (pp. 89–105). Washington, DC: Hemisphere.

Harris, C. L., Gleason, J. B., & Ayçiçeği, A. 2005. "When is a first language more emotional? Psychophysiological evidence from bilingual speakers". In A. Pavlenko (Ed.), *Bilingual minds: Emotional experience, expression, and representation* (pp. 257–283). Clevedon, UK: Multilingual Matters.

Harris, C.L. (2004). Bilingual speakers in the lab: psychophysiological measures of emotional reactivity. *Journal of Multilingual and Multicultural Development, 25,* 223–247.

Harris, C.L., Ayçiçeği, A. & Gleason, J.B. (2003). Taboo words and reprimands elicit greater autonomic reactivity in a first than in a second language. *Applied Psycholinguistics, 24,* 561-579.

Herbert, C., Junghofer, M., & Kissler, J. (2008). Event related potentials to emotional adjectives during reading. *Psychophysiology, 45,* 487-498.

Janschewitz, K. (2008). Taboo, emotionally valenced, and emotionally neutral word norms. *Behavior Research Methods, 40(4)*, 1065-1074.

Jay, T., Caldwell-Harris, C., & King, K. (2008). Recalling taboo and nontaboo words. *American Journal of Psychology, 121,* 83–103.

Kissler, J., Assadollahi, R., & Herbert, C. (2006). Emotional and semantic networks in visual word processing: Insights from ERP studies. *Progress in Brain Research, 156*, 147-183.

Kissler, J., Herbert, C., Peyk, P., & Junghofer, M. (2007). Buzzwords : Early Cortical Responses to Emotional Words During Reading. *Psychological Science, 18,* 475-480.

Kissler, J., Herbert, C., Winkler, I., & Junghofer, M. (2009). Emotion and attention in visual word processing – An ERP study. *Biological Psychology, 80,* 75-83.

MacKay, D. G., Shafto, M., Taylor, J. K., Marian, D. E., Abrams, L., & Dyer, J. R. (2004). Relations between emotion, memory, and attention: Evidence from taboo Stroop, lexical decision, and immediate memory tasks. *Memory & Cognition, 32,* 474-488.

Pavlenko, Aneta. 2007. *Emotions and Multilingualism.* New York: Cambridge University Press.

Perea, M., Carreiras, M., & Grainger, J. (2004). Blocking by word frequency and neighborhood density in visual word recognition: A task-specific response criteria account. *Memory & Cognition, 32*, 1090-1102.

Sutton, T. M., Altarriba, J., Gianico, J. L., & Basnight-Brown, D. M. (2007). The automatic access of emotion: Emotional Stroop effects in Spanish-English bilingual speakers. *Cognition and Emotion, 21,* 1077-1090.

Appendix A

Critical Items

VA

amazing	glow
award	harmony
caressed	honored
champion	intelligence
cheer	interest
clever	lavish
comforted	marry
cuddle	plentiful
daring	praise
delicious	prize
devotion	proud
elation	reward
embraced	save
encourage	sunny
engagement	thrilled
excels	unique
excited	upbeat
fantastic	uplifted
fireworks	valuable
glamorous	victory

Va

admires**	loyal**
affection	lullaby
asset	mature
awoke	millionaire
benefactor	peace
bonus	plays
children*	pristine**
chuckled*	purify
coming	rainbow
compensation	reasonable
cook	recommended
cookies	rejoiced
cozy	relax
decent	right
food	selfless**
gentle	sensible
greet*	serenity
joke	sparkling
kind**	studious
knowledge	truest
leaves*	tulips
likable	wonders

* = Experiment 1 only
** = Experiment 2 only

vA	
anger	anger
anguish	anguish
arrest	arrest
bombs	bombs
cheated	cheated
creepy	creepy
crisis	crisis
critical	critical
damaging	damaging
dangerous	dangerous
death	death
despair	despair
disaster	disaster
dying	dying
edgy	edgy
failure	failure
fake	fake
grotesque	grotesque
horrendous	horrendous
humiliated	humiliated

va	
addict	horrors
alone	ignorance
anxiety	inappropriate
blind	infected
boring	infuriated
canonical	moldy
cramped	noose
criminal	pathetic
criticism	poop
cry	problems
debt	queasy
decay	slob
denied	slurs
detention	sprain
disliked	stupid
dump	tacky
fool	trash
foul	ugly
gangster	useless
garbage	virus

Neutral

absurd
adequate
alien
attack
charged
chiseled
complex
cringe
crunchy
denounced
expensive
factual
forged
galloped
heroine
hidden
larger
mystery
nuzzled
offbeat
penalty
polite
recognized
register
resents
ridiculous
scalded
sharp
sketched
splashed
struck
swayed
swears
Swedish
troubling
unpredictable
unusual
worried
worship
wrong

Taboo

anus
ass
asshole
bastard
bitches
boner
boobs
breasts
bullshit
butt
clitoris
cocks
crappy
cunt
damn
dicks
dildo
douche
dyke
faggots
fellatio
fuck
fucking
goddamn
hell
hoe
homo
hooker
humping
nipples
pecker
pee
penis
piss
prick
pussy
sex
shitty
sluts
snatch
suck
testicles
tits
twat
vagina
wanker
whores