The mental representation and social aspect of expressives

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Cognition and Emotion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>CEM-FA 12.20.R2</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Full Article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>n/a</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Donahoo, Stanley; University of Arizona, Linguistics Tzuyin Lai, Vicky; University of Arizona, Psychology</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Swearing, ERP, Emotion, Language, Taboo</td>
</tr>
</tbody>
</table>
The mental representation and social aspect of expressives

Author Names and affiliations:

Stanley A. Donahoo\textsuperscript{a,c}, Vicky Tzuyin Lai\textsuperscript{b,c}

\textsuperscript{a}Department of Linguistics, University of Arizona
\textsuperscript{b}Department of Psychology, University of Arizona
\textsuperscript{c}Cognitive Science Program, University of Arizona

Data Availability Statement: Data available on request from the authors.
Abstract
Despite increased focus on emotional language, research lacks for the most emotional language: Swearing. We used event-related potentials (ERPs) to investigate whether swear words have content distinct from function words, and if so, whether this content is emotional or social in nature. Stimuli included swear (e.g. shit, damn), negative but non-swear (e.g. kill, sick), open-class neutral (e.g. wood, lend), and closed-class neutral words (e.g. while, whom). Behaviourally, swears were recognised slower than valence- and arousal-matched negative words, meaning that there is more to the expressive dimension than merely a heightened emotional state. In ERPs, both swears and negative words elicited a larger positivity (250-550 ms) than open-class neutral words. Later, swears elicited a larger late positivity (550-750 ms) than negative words. We associate the earlier positivity with attention due to negative valence, and the later positivity with pragmatics due to social taboone. Our findings suggest a view in which expressives are not merely function words or emotional words. Rather, expressives are emotionally and socially significant. Swears are more than what is indicated by valence or arousal alone.

Keywords: Swearing, ERP, Emotion, Language, Taboo

Words: 7,942
1. Introduction

Swear words are words such as *damn, hell*, and *shit*, and are often perceived as both negative and highly arousing. How are swear words understood? Traditionally, swears have been considered categorically as closed-class, function words, much like prepositions (Leech et al., 1982; Segalowitz & Lane, 2000). However, swear words differ from function words in that they give rise to physical effects. For instance, swearing increases tolerance for physical pain (Stephens et al., 2009), and elicits stronger skin conductance responses than neutral words (Bowers & Pleydell-Pearce, 2011). Recent theories consistent with such a lack-of-content view propose that swear words have no propositional content themselves. Rather, they exert influence on their neighbouring content words (Gutzmann, 2019; Potts, 2005b, 2007). Further, this lack of propositional content allows for great flexibility as to where the swear is interpreted with respect to the surrounding lexical context (Frazier et al., 2015, 2017). Another rational way to understand swear words is to treat them as emotional words, because swear words are oftentimes perceived as both negative and arousing. Following this line of reasoning, is the category of swear words a special kind of negative words? In this study, we propose that swears are neither just function words nor emotional words and that there is some use-conditional, social content encoded in swear words. We use event-related potentials (ERPs) to investigate whether swear words have content distinct from function words, and if so, whether this content is emotional or social in nature.

According to linguistic theories that view swear words as lacking propositional content, swearing is studied as expressive content (Gutzmann, 2019; Potts, 2007). Expressives are thus considered not-at-issue content, which is difficult to isolate from the utterance in which it occurs (Simons et al., 2010). For example, an infelicitous follow-up to *The damn dog is on the couch* would be something like, *That’s not true, Fido is a lovely dog*, demonstrating the difficulty an interlocutor has in trying to isolate the expressive content on
its own. Given that the expressive content of the swear word *damn* is not what is at-issue in an utterance, it has been argued that the expressive content represents a dimension of meaning which is separate from the rest of the descriptive, truth-conditional content of the utterance (Gutzmann, 2013; Potts, 2005a, 2007; Anderbois, Brasoveanu, & Henderson, 2015; Barker, Bernardi, & Shan, 2010). Evidence in support of an additional dimension of meaning are provided by both clinical and neurotypical populations.

Spontaneous usage of expressive content is an infamous trait of Tourette syndrome (Robertson, 2000; Van Lancker & Cummings, 1999). In healthy populations, Dörre and colleagues (Dörre et al., 2018) examined not-at-issue content using German modal particles and found processing differences compared to their at-issue counterparts, suggesting differences in meaning representations.

In emotion research, many view swear words as being negatively valenced and highly arousing (Reilly et al., 2020). Emotional language researchers typically assume a two-dimensional model, comprised of valence and arousal (Citron, 2012; Citron et al., 2013; Russell, 1980; Lang, Bradley, & Cuthbert, 1998). Valence represents a value measure where higher scores represent a word that could be considered good or positive, whereas lower values represent a word considered bad or negative. Arousal is an excitement measure, where highly arousing words produce feelings of stimulation or excitement and low arousal words produce feelings of relaxation or calmness (Russell & Barrett, 1999; Scott et al., 2018).

In what follows we briefly review some of the behavioural and electrophysiological literature on emotional words relevant to swears (Citron, 2012).

Behavioural studies have found faster and more accurate recognition for words with any emotional valence, be it negative or positive, in comparison with neutral words (Citron et al., 2013; Kousta et al., 2009). This processing advantage is theorised to represent the relevance of emotional stimuli in both survival generally and goal attainment specifically.
While the directionality of the effects of word valence are mixed, (Algom et al., 2004; Kanske & Kotz, 2007; Knickerbocker et al., 2015; Kuperman et al., 2014; Scott et al., 2014; Wentura et al., 2000), overall, researchers have taken a theoretical stance which favours a processing advantage for negative words as a reflection of survival often coinciding with fleeing from negative situations (Fox et al., 2001; Knickerbocker et al., 2015). Generally, emotional stimuli, regardless of valence directionality, capture attention because they are motivationally significant (Lang et al., 1990, 1997). More recently, it has been argued that prior studies on emotional language processing conflated one or more linguistic variables (e.g. word frequency), and that, when all known variables are accounted for, we are left with a U-shape model where more extreme valence values across the spectrum speed up lexical decision times (Kuperman et al., 2014). Further, while arousal has been shown to facilitate lexical decision times when valence is held constant (Hofmann et al., 2009), valence, rather than arousal, has been demonstrated to be a more robust predictor in modelling factors of lexical access (Bradley & Lang, 2000; Kousta et al., 2009; Kuperman et al., 2014; Vinson et al., 2014). Specifically, valence is a significant predictor of lexical decision response latencies, even when other factors, including arousal, are included in regression models (Kousta et al., 2009). Finally, valence, unlike arousal, appears to remain a modulating factor in tasks beyond the word level (Bayer et al., 2010).

Electrophysiological studies, with their superior temporal resolution, inform us that the emotional content of a word is activated rapidly. Event Related Potential (ERP) studies of emotional words have identified at least two neural correlates (Citron, 2012). One is the early posterior negativity (EPN), which peaks roughly between 200 and 300 ms after stimulus onset, and is larger for both positive and negative valenced words than for neutral words (Herbert, Junghofer, & Kissler, 2008). It has been suggested that the EPN reflects automatic, implicit processing of emotion (Palazova et al., 2011). However, studies which reported an
EPN tend to employ tasks distinct from natural reading, like multiple repetitions of single words (Kissler et al., 2007) or counting the number of times a word from a grammatical class occurs (Kissler et al., 2009). A second neural correlate robustly associated with emotional word processing is the late positive component (LPC), which peaks roughly between 500 and 750 ms (Bayer et al., 2010, 2012; Bayer & Schacht, 2014; Yao et al., 2016). The LPC is an index of a more sustained, explicit, processing of emotional content; or, put another way, the LPC indexes higher-order stimulus evaluation (Schacht & Sommer, 2009). Negative words have been shown to increase the amplitude of the LPC (Kanske & Kotz, 2007; L. Wang & Bastiaansen, 2014). Collapsing over positive and negative words, emotional words elicited a larger LPC relative to neutral words, a finding in parallel with behavioural studies, where valence tends to facilitate lexical access (Citron et al., 2013; Kousta et al., 2009). In general, more valenced words produce larger ERP amplitudes, which is associated with emotional processing (Citron et al., 2013; Kanske & Kotz, 2007).

If the not-at-issue content in swear words is primarily emotional, we would expect to find the ERP correlates discussed in the above-mentioned literature. However, here we suggest that the two dimensions of valence and arousal may not fully capture the use-conditional content of swear words. In addition to emotionality, there are at least three possibilities; the content is: (1) a pragmatic speech act, (2) tabooness, (3) a socially threatening function. Frazier, Dillon, and Clifton (2015, 2017) argue for pragmatically treating expressives as independent speech acts, independent of the speech act conveyed by the rest of the utterance. As a result, swears thus have flexibility as to where they are applied to the rest of the at-issue content. Frazier and colleagues found evidence for this claim using offline sentence ratings. Participants read sentences such as The policeman gave me a damn ticket and rated whether the speaker felt negative towards the ticket (object reading), the policeman (subject reading), or the entire situation (situation reading) itself. Even when the
expressive was in an object position, participants still preferred the situation response nearly 42% of the time. Similarly, damn presented in subject position yielded a situation interpretation in 39% of ratings and damn in situation position (e.g. Damn. The policeman gave me a ticket.) yielded a situation interpretation in only 63% of ratings, empirically demonstrating the flexibility of not-at-issue swear.

Tabooness is another potential factor that may drive swear word recognition. Taboo language generally produces slower reaction times across a broad variety of tasks, including modified Stroop (Eilola & Havelka, 2011; Guillet & Arndt, 2009; Mackay et al., 2004; Siegrist, 1995), attentional blink (Mathewson et al., 2008), and picture-word interference (Dhooge & Hartsuiker, 2011). This slowdown is conversely associated with the taboo-superiority effect: the phenomenon that memory is better for taboo words over neutral words in a variety of experimental paradigms, for example in spontaneous recall tasks (Hadley & MacKay, 2006; Mathewson et al., 2008). Better recall has also been demonstrated for taboo words in word lists (Buchanan et al., 2006; Grosser & Walsh, 1966), and in repetition priming (Thomas & LaBar, 2005). Madan, Shafer, Chan, & Singhal, (2017) reported that factors other than valence, such as word frequency and tabooness, are the better predictors of lexical decision latencies. It should be mentioned however, that stimuli in Madan et al. (2017), while including words that could be considered offensive (e.g. skank), do not directly feature any words that are commonly considered swears (e.g. damn, shit, hell).

Lastly, social threat has been associated with swear words. One study explored the idea that social threat and physical threat are distinct (Wabnitz et al., 2012). Participants passively read socially threatening negative words (swear words), physically threatening negative words (e.g. bomb, explosion, wound), neutral, and positive words, while monitoring for a dot presented in 15% of trials. The socially threatening negative words (swear words) elicited an enhanced P1 (100-140 ms) relative to physically threatening negative words and
positive words. The authors interpreted the P1 as a reflection of a fast-forward mechanism for evolutionarily relevant cues. In their so-called N400 time window (450-580 ms), negative and positive words elicited a larger negativity effect relative to neutral words, with no difference between the positive and negative words. The authors interpreted the N400-like effect as evidence that swear words receive no special semantic consideration relative to other emotionally valenced words, at least not in a typical population. In people with social anxiety disorder, swears did entertain semantic consideration in the N400 time window (Wabnitz et al., 2016). A related study (Klein et al., 2019) reported LPC differences for socially threatening and physically threatening words in adolescents with PTSD, but not in healthy controls.

The present study uses event-related potentials (ERPs) to investigate the nature of the content, if any, in swear words. Participants engaged in a lexical decision task while EEG was simultaneously recorded on the scalp. We compared the response times and ERPs for swear words, negatively valenced but non-swear words, neutral open class words, and neutral closed-class function words. If swears are not as content heavy in their being not-at-issue, they should behave similarly to function words. If swears have content and if the content is emotional in nature, then swears should behave more like negative words. If swears have content and if the content is something more than emotional in nature, then swears should be distinct from negative words.

2. Material and methods
2.1. Participants
Thirty-two right-handed native English speakers (15 female; age range 18-23) participated in the experiment for course or extra credit. According to power analyses using G*Power software (Faul et al., 2009), this was a sufficient sample size using a conventional threshold of $\alpha = .05$ for Type I error, a power of .80, and a medium effect size (partial
η² = .06). All participants had normal or corrected-to-normal vision, and had no known language or neurological disorder. All gave informed consent before participation. For institutional review board purposes, participants were verbally informed that stimuli presented during the experiment included words that could be considered offensive; but, in order to avoid any sort of priming effects, no examples of such words were provided before the experiment began.

2.2. Stimuli

The stimuli consisted of 28 swear words (e.g. *shit, damn*), 28 negatively valenced but non-swear words (*kill, sick*), 28 open-class neutral words (e.g. *wood, lend*), 28 closed-class neutral words (e.g. *while, whom*), and 112 pseudowords. The number of words is largely constrained by the number of words considered swears. The words were generated by native-speaking research assistants with reference to the swear literature (Jay, 1992; Wang, Chen, Thirunarayan, & Sheth, 2014) and the literature on closed class words (Craig & Kinney, 2009). Two web-based norming surveys verified the manipulations of valence and tabooiness. Thirty-nine native-speaking participants rated the words one word at a time on a 9-point Likert scale. In the valence norming (“how positive or negative a word is”), 1 was negative and 9 was positive. In the norming for tabooiness (“how offensive the word is to people in general”), 1 was not at all offensive and 9 was very offensive. Arousal values were taken from existing corpus work (Warriner et al., 2013), supplemented by the web-based norming survey for the closed-class words. Based on the results of the norming survey, one stimulus item initially coded as a swear word, *christ*, was removed from the analysis, as participants consistently rated the word as relatively positive (6.19) and not taboo (4.75).

Table 1 summarises the word properties. Both swears and negative words were rated more negative than neutral words, (ts > 12, ps < .001), and swears and negative words were similarly negative (n.s.), and similarly arousing (n.s.). Swears were rated more taboo than
negative words ($t(43) = 6.74\ p < .001$). In addition, the conditions were matched in terms of log frequency (Brysbaert & New, 2009), word length, and number of morphemes, phonemes, and syllables (Balota et al., 2007). There was no significant difference between these conditions.

The pseudowords were constructed using the ARC Pseudoword Database (Rastle et al., 2002), selecting only strings with orthographically existing onsets, and using an average range of letter length (min: 4, max: 6), as well as orthographic (min: 4, max: 6) and phonological (min: 12, max: 14) neighbourhood sizes that matched the target words in the task.

2.3. Procedure

Preparation for the EEG recording lasted 30 minutes, in which participants were measured and fitted with the electrode cap (ActiCap, Brain Products, GmbH). During the setup, participants provided general language and demographic information and were briefed as to the nature of the EEG technique. Participants then entered a dimly lit and sound attenuated room. They were seated in a comfortable chair at a desk which featured a computer screen 70-80 cm away from their eyes. The stimuli were presented visually in white font (Font: Lucida Console; Font size: 20) against a black background via the E-prime 3.0 software (Psychology Software Tools, Inc.).

Figure 1 shows an example of a trial. In each trial, a central fixation cross was initially presented for 500ms, followed by a blank screen with a random duration of 500-1500 ms. The stimulus (either a legal letter string or illegal letter string) then appeared for a duration determined by the number of letters in the string, ranging from 248-400 ms (37 ms/letter + 100 ms, max duration of 400 ms). Participants were instructed to decide whether the letter string was a word or not as quickly and as accurately as possible by pressing one of two buttons on a response pad. The order of the button press was counterbalanced with
participant number. A trial timed out after 2000 ms, but this duration proved adequate to garner a participant response.

There were 6 practice trials to familiarize participants with the task. Participants were advised to refrain from moving or blinking during the trials, but encouraged to rest their eyes at break intervals, occurring after every 8 trials. Each item was presented once, in one block with a randomised presentation order. We decided against repeating items, as there is evidence that emotional words in particular are susceptible to repetition priming effects, both immediate and delayed (D. Zhang et al., 2019).

2.4. EEG acquisition and pre-processing

Continuous EEG was recorded from 64 surface active electrodes using the actiCHamp system (Brain Products, GmbH). Data were measured with respect to a vertex reference, with the sampling rate of 500 Hz. Electrode impedances were below 10 kΩ throughout.

Data was analysed with Brain Vision Analyzer 2.0. Data was low-pass filtered at 30 Hz with a notch filter at 60 Hz to filter out line noise. Blinks were detected and corrected offline via ocular artefact correction independent component analysis (ICA) using the Infomax Restricted ICA algorithm. Artefacts were detected semi-automatically with Brain Vision. Trials contaminated with artefacts due to body movements or peak deflection exceeding ±100 µV were rejected. Individual bad channels were replaced through interpolation of the average of the surrounding channels (Brain Vision User Manual) on a participant-by-participant basis. Remaining artefacts were reduced by individual channel-interpolation or the epochs were rejected. 1.5% of the channels were replaced across the experiment. Data were re-referenced to the average of the left and right mastoids, and segmented from 200 ms before the stimulus onset to 750 ms after, with baseline correction from -200 ms to 0 ms preceding the onset of the target letter string. Finally, the ERP data were averaged across participants for each condition.
3. Results

3.1. Behavioural

Mean behavioural accuracy and reaction times results for trials with correct responses are displayed in Figures 2 and 3. Separate Repeated-Measures ANOVAs were conducted to compare the effect of word condition (Swear, Negative, Neutral, Closed, Pseudoword) on accuracy and on reaction times, respectively. Greenhouse-Geisser corrected $p$ values are reported when appropriate for all tests involving factors with more than two levels. In addition, False Discovery Rate corrections were applied (Benjamini & Hochberg, 1995). In terms of accuracy, there was a significant effect for condition, $F(4,112)=20.38, p < .001, \eta^2_G = .259$. Crucially, there were no differences for swear and neutral, $F(1,28)=1.70, p = .303$; negative and neutral, $F(1,28)=.15, p = .697$; nor swear and negative, $F(1,28)=1.29, p = .319$. There were significant differences between pseudoword and neutral $F(1,28)=21.83, p < .002, \eta^2_G = .199$; closed and neutral, $F(1,28)=53.65, p < .002, \eta^2_G = .312$; and closed and swear, $F(1,28)=26.48, p < .002, \eta^2_G = .236$.

In terms of reaction time, there was a significant effect for condition, $F(4,112)=47.58, p < .001, \eta^2_G = .139$. Pairwise comparisons revealed that there were significant differences between pseudoword and neutral, $F(1,28)=132.59, p < .001, \eta^2_G = .198$; swear and neutral, $F(1,28)=18.03, p < .001, \eta^2_G = .030$; swear and negative, $F(1,28)=18.5, p < .001, \eta^2_G = .056$; closed and neutral, $F(1,28)=30.68, p < .001, \eta^2_G = .054$. Negative compared to neutral was not significant, $F(1,28)=1.69, p = .204$. Closed compared to swear was not significant, $F(1,28)=1.90, p = .204$.

3.2. ERP

The grand averaged ERPs are displayed in Figure 4, and the scalp distributions of the effects, in Figure 5. Average trial counts for each condition were swear: 26.78; negative: 25.75; neutral: 25.97; closed: 26.38; pseudoword: 101.53. Descriptively, the overall amplitude for swear was greater than the other conditions at early (250-550) and later (550-
750) time windows: swear, $M = 2.74 \mu V$; negative, $M = 2.12 \mu V$; neutral, $M = 1.17 \mu V$; closed, $M = 1.33 \mu V$; pseudoword, $M = -.55 \mu V$. 550-750: swear, $M = 5.64 \mu V$; negative, $M = 3.41 \mu V$; neutral, $M = 2.85 \mu V$; closed, $M = 3.87 \mu V$; pseudoword, $M = 1.95 \mu V$.

Visual inspection indicated that all waveforms showed early components of N1 and P2, which reflect visual processing and reassure data quality. The conditions diverged as early as ~250 ms. Negative words elicited a larger positivity than neutral words from 250-550 ms, centrally and parietally distributed. Swears also elicited a larger positivity than neutral content words from 250-550 ms, widespread in all locations, but showed more positive amplitudes than negative words from 550-750 ms. Pseudowords elicited a larger negativity than neutral open-class words starting at 250 ms, identified as N400. There was no EPN (See supplementary materials).

Mean amplitudes were exported from two time windows of 250-550 ms and 550-750 ms at the frontal group (Afz, F1, F2, Fz), central group (FCz, C1, C2, Cz), and the posterior group (CPz, P1, P2, Pz) along the midline. Greenhouse-Geisser corrected $p$ values are reported when appropriate for all tests involving factors with more than two levels. False Discovery Rate corrections were also applied. The omnibus Repeated-Measures ANOVA model of 5 condition (Swear, Negative, Neutral, Closed, Pseudoword) x 3 location (Front, Central, Posterior) x 2 time (Early, Late) showed an interaction of condition x time x location, $F(8,248)=5.19, p < .001, \eta^2_G = .001$, and an interaction of condition x time $F(4,124)=5.70, p < .001, \eta^2_G = .004$. These interactions suggest that the patterns of results in the two time windows differed. Thus separate Repeated-Measures ANOVAs of condition x location were conducted within each time window.

### 3.2.1. 250-550 ms

There was a condition x location interaction, $F(8,248)=3.36, p = .010, \eta^2_G = .002$, a main effect of condition $F(4,124)=14.41, p < .001, \eta^2_G = .067$, and a main effect of
location $F(2,62) = 40.48, p < .001, \eta^2_G = .200$. Pseudoword elicited a larger and widespread N400 than neutral words (Frontal: $F(1,31) = 18.40, p < .004, \eta^2_G = .060$; Central: $F(1,31) = 23.08, p < .004, \eta^2_G = .068$; Posterior: $F(1,31) = 9.57, p = .013, \eta^2_G = .028$), consistent with past ERP findings for pseudowords (Palazova et al., 2011; Schacht & Sommer, 2009a). Figure 6 displays relevant ERPs and scalp topographies for these two conditions.

Comparing negative and neutral words, negative words elicited a larger positivity than neutral words, centrally and posteriorly distributed (Frontal: $F(1,31) = 3.43, p = .074, \eta^2_G = .009$; Central: $F(1,31) = 5.35, p = .033, \eta^2_G = .018$; Posterior: $F(1,31) = 7.66, p = .019, \eta^2_G = .020$).

Comparing swear and neutral words, swear words elicited a larger positivity than neutral words, at all locations (Frontal: $F(1,31) = 9.04, p = .013, \eta^2_G = .037$; Central: $F(1,31) = 5.89, p = .027, \eta^2_G = .030$; Posterior: $F(1,31) = 6.78, p = .020, \eta^2_G = .025$).

Comparing swear and negative words, the two behaved similarly in this time frame. That is, swears elicited a larger positivity than negative words in the frontal group only (Frontal: $F(1,31) = 4.28, p = .050, \eta^2_G = .011$; Central: $F< < 1$; Posterior: $F< < 1$).

Closed class words behaved similarly to neutral (content) words in this window (all $Fs < 1$). But closed class words behaved differently compared to swears, with swear eliciting a larger positivity (Frontal: $F(1,31) = 12.46, p = .004, \eta^2_G = .032$; Central: $F(1,31) = 7.31, p = .019, \eta^2_G = .020$; Posterior: $F(1,31) = 7.1, p = .019, \eta^2_G = .014$).

3.2.2. 550-750 ms

There was a main effect of condition, $F(4,124) = 11.41, p < .001, \eta^2_G = .053$, but condition did not interact with location, $F< < 1$. Thus in the following analyses we combined
all electrode locations. Pseudoword elicited a marginally larger negativity than neutral words 
\(F(1,31) = 4.17, p = .07, \eta^2_G = .013\), likely a continuation of the N400.

Swears elicited a larger positivity than neutral words, \(F(1,31) = 17.54, p < .003, \eta^2_G = .078\). Swear also elicited a larger positivity than negative words, \(F(1,31) = 12.75, p = .003, \eta^2_G = .049\). Negative and neutral words do not differ here \(F(1,31) = 1.74, p = .197\).

Closed class words behaved similarly to neutral (content) words in this time window, 
\(F(1,31) = 2.76, p = .126\). But closed class words behaved differently compared to swear words, with swear words eliciting a larger positivity than the closed \(F(1,31) = 9.96, p = .008, \eta^2_G = .026\).

Further, to explore task difficulty on ERP responses, using mean reaction times for hits in the respective emotion conditions as a within-subject covariate, the ANCOVA indicated that in the late positivity window (550-750 ms), the effect of swear words still differs from the effect of negative words \(F(1,62) = 7.97, p = .006\). This suggests that after taking task difficulty (approximated by reaction times) into consideration, swear words are processed differently from the negative words in the late time window. In the early positivity window (250-550 ms), the effect of swear words does not differ from the effect of negative words \(F(1,62) = .817, p = .369\), suggesting that after taking task difficulty into consideration, swear words and negative words are processed similarly in this earlier positivity time window.

4. Discussion

The present study investigated whether swear words have an additional dimension of meaning, and if so, whether such meaning is functional, emotional, or social in nature. Participants read swear words, negatively valenced but non-swear words, neutral open-class words, neutral closed-class words, and pseudowords. They made word/non-word decisions...
while their EEG was recorded. The behavioural results showed that, while participants were equally as accurate to respond to swears, negative, and neutral words, swear words were recognised more slowly than neutral open-class words, and neutral open-class words were recognised more slowly than negative words. Swears were recognized as quickly as the closed-class neutral words. For ERPs, both swears and negative words elicited larger positivity effects in the 250-550 ms time window relative to neutral words. In the 550-750 ms time window, the positivity effect in swears continued but the positivity effect in the negative words condition converged with the neutral. Swears were also distinct from closed-class neutral words in both time windows, in that swears elicited larger positivity effects than closed-class neutral words.

4.1. Behavioural results

That swear words were recognised more slowly than neutral open-class words could be interpreted as a taboo superiority effect, and that neutral open-class words were recognised more slowly than negative words could be viewed as an example of well-attested negative bias (Ito et al., 1998; L. Wang & Bastiaansen, 2014). Recent behavioural work in speech production has argued that taboo effects are pre-lexical (Hansen et al., 2017). However, our results suggest that taboo effects emerge later, around 250 ms—at which lexical access already occurred (Dien, 2009; Kim & Lai, 2012). Looking at the issue of when emotional meaning affects word processing specifically, Palazova and colleagues used lexical decision simultaneously with EEG to conclude that emotionality has a post-lexical focus (Palazova et al., 2011). While there were behavioural differences for open-class and closed-class neutral words, with the closed-class words being identified more slowly and less accurately, there were no ERP differences in the two, suggesting that the reaction time differences are attributable to overall familiarity with the two conditions.
In our opinion, ERPs and behaviours are measuring different aspects and timings of the same process, and therefore behaviours might not always explain ERPs. Considering reaction times and ERPs for swears and negative words together, a relatively consistent picture emerges: The negative valence (present in both swears and negative words), or rather the attention allocation to negative valence is processed relatively earlier, whereas the social taboo aspect (present in swears and not in negative words) is processed later, after 550 ms.

4.2. Swears and early effects

Our results indicate that swears in isolation were perceived similarly to negative words initially until 550 ms, and then continued to be processed for their pragmatic content. Both swears and negative words elicited larger ERP positivity effects relative to neutral words in the 250-550 ms window. We are confident in associating the positivity effect here with negative valence, based on past literature (Hinojosa, Albert, López-Martín, & Carretié, 2014; Kanske & Kotz, 2007; L. Wang & Bastiaansen, 2014; Zhang, Ge, Kang, Guo, & Peng, 2018, but see Bayer & Schacht, 2014). Note that compared to those studies, our positivity effect time window 250-550 ms seems earlier, e.g., 400-650 ms in Hinojosa et al. (2014). We included an earlier portion because we considered effects in all conditions, including the pseudoword effect which started at 250 ms. Such inclusion mathematically counteracted statistical significance, yet we still found significance, pointing to the robustness of the positivity effect. An additional post-hoc analysis using the 400-600 ms window verified the significance of the larger positivity effect for negative than for neutral words ($F(1,31)= 7.88, p = .009, \eta^2_G = .017$).

We attribute the positivity effect in the 250-550 ms window to attention allocation (Hinojosa et al., 2014). According to the Global Resource Theory (Mackay et al., 2004), attentional capacities are finite. Emotional words demand more attentional resources and incur a processing disadvantage, which possibly accounts for the delay we observed in lexical
decisions times for participants responding to the swear condition. A related approach, automatic vigilance (Erdelyi, 1974; Pratto & John, 1991; Wentura et al., 2000), suggests that negative stimuli engage attention longer than neutral stimuli because humans preferentially attend to negative stimuli, though it remains unclear at what stage of processing the effect arises (Kuperman et al., 2014). Task type and relevance also appear to play a role in the processing of negative compared to neutral stimuli (Trauer et al., 2012, 2015).

We did not find any sort of EPN for the swear words or the negative valence words, replicating some previous work (M. Zhang et al., 2018; M. Zhang & Guo, 2014). Those studies reporting an EPN typically found it for positive words or pictures (Hinojosa et al., 2009; Schacht & Sommer, 2009b), and swears tend to be used in negative contexts. Future work should examine whether swears used in a positive context (e.g. sheit is damn prettyfucking awesome) would give rise to an EPN.

4.3. Swears and the late positivity

Our primary finding is that there is more to swearing than negative affect, as reflected by the late positivity effect in the 550-750 ms window which was more enhanced for swears than for negative words. There are a few possible interpretations for this. The first, and perhaps most parsimonious, interpretation is that the late positivity was merely a continuation of the positivity effect (or a more extreme version of the effects). The effect was prolonged in swears because some function in swears acted to amplify their emotional properties and made the perceived emotionality more robust than what those properties actually were. This interpretation is partially consistent with the not-at-issue view of swears. We say only partially because the not-at-issue view suggests that swears impact words surrounding them, and here as a single word study, there were no words neighbouring the swears.

A more likely interpretation is that the late positivity effect is a real effect in its own right. Under this scenario, swears contribute their own dimension of meaning and the late
positivity could reflect variables inherent to swears, such as: (1) a pragmatic speech act, (2) social threat, (3) tabooness. Ultimately, we conclude that the tabooness element unique to swears is the best candidate, but we take each of these in turn.

First, the late positivity could reflect an extra step associated with processing a speech act. This would support the view in which expressives represent their own distinct speech act (Frazier et al., 2015, 2017). Theoretically, there are several factors to consider regarding any one speech act (Bach, 1998). These extra steps come with processing costs, some of which have been borne out experimentally for both direct and indirect speech acts (Coulson & Lovett, 2010; Gísladóttir, Bögels, & Levinson, 2018; Gísladóttir, Chwilla, & Levinson, 2015). It is difficult to compare these ERP findings for multi-word speech acts to our lexical decision task. Gísladóttir and colleagues reported a late negativity for speech acts in which an interlocuter makes an indirect offer (by saying I have a credit card as a response to I don’t have any money to pay for the ticket) as opposed to a direct answer (I have a credit card as a response to How are you going to pay for the ticket?). Coulson and Lovett (2010) reported LPC effects for differences in literal statements and indirect requests (My soup is too cold to eat, said to either a husband or a restaurant waiter), with a larger LPC for indirect requests. However, this positivity only emerged for words later in the request, much later than we see for words presented in isolation. Still, negative and neutral words alone in isolation do not represent a speech act in the way that a swear word might; for example, uttering shit after spilling coffee on a computer desk can invoke a hearer to move the keyboard. However, it remains an open question whether a swear word presented in a lexical decision task can be construed as conveying communicative meaning in the way that a speech act typically does.

Second, the late positivity could reflect the processing of social threat, an idea postulated for differences observed for taboo words in the two extant ERP studies directly involving swearing (Wabnitz et al., 2012, 2016). Wabnitz and colleagues argued that swear
words receive no special semantic consideration relative to other emotionally valenced words while our findings suggest otherwise. One likely explanation for the differences in our ERPs can be attributed to task differences. Wabnitz and colleagues relied on passive reading while our study required that participants made an active decision about each letter string. Such task differences are known to affect lexical processing for emotional words (Kissler et al., 2009). Additionally, Wabnitz and colleagues employed a block design, meaning that participants saw the words in each condition six times. Such repetition can significantly alter ERPs (Van Petten et al., 1991). Finally, there is a potential confound in the design of both of these studies regarding the label of swear word towards the stimuli used, however. Because the research question was interested in navigating social threat, the swear stimuli used were person-oriented, thus functioning more like an insult than a swear word specifically. Categorically, there is a difference in fuck or fucking compared to motherfucker (Blakemore, 2015). The former are swears, whereas the latter is also an insult. All of the stimuli used by Wabnitz and colleagues are person-oriented and interpreted as a noun. It remains an open question as to whether these related but distinct conceptual categories are processed and represented in similar fashion. In this vein, Klein and colleagues (2019) used the same stimuli as Wabnitz and colleagues and found no difference in insults and physically threatening words in healthy adolescents, whereas they did find differences for adolescents with PTSD. In general, all of these words bear some social offense, but there are linguistic distinctions. It is clear however, that expressives do receive special semantic consideration relative to other words that are equally as emotionally valenced.

Third, the late positivity could reflect the processing of an additional dimension, tabooness. That is, even though damn and death are both negatively valenced, only damn has the additional use-conditional taboo dimension. Swearing is not merely the sum of extreme values of valence and arousal. Instead, these words offer their own element of affect, namely
tabooness, and contribute their own dimension of meaning, both of which are distinct from other words that are just as extreme in their valence values. This interpretation is consistent with recent experimental work where tabooness demands more attention in sentential contexts (Christianson et al., 2017). In our lexical decision task, relative to both negative and neutral words, swears elicited longer reaction times, a finding consistent with prior studies involving a variety of tasks (Dhooge & Hartsuiker, 2011; Eilola & Havelka, 2011; Guillet & Arndt, 2009; Mathewson et al., 2008). Regarding our ERP findings, it is a challenge to situate the positivity effect for swears relative to negative and neutral words, given that so few studies have investigated these specific aspects of emotional language. Still, previous work shows that negative words increase the amplitude of the LPC (Kanske & Kotz, 2007; Wang & Bastiaansen, 2014).

The limitations of this study are as follows: (1) The trial counts are relatively low (comparatively lower (~26 each condition), which is due to the fact that there are not that many swear words to begin with and that we had to reject trials contaminated by artefacts. However, the data are still relatively reliable because first, within this dataset, early visual components N1 and P2 are clearly present, which verifies data quality in general. Second, beyond this dataset, other studies have used comparable trial counts, e.g., 20 in Hinojosa et al. (2009), 32 in Kanske et al. (2011), and 20 in Yao et al. (2016). Thus, while we should keep this limitation in mind, the observed effects are relatively reliable. Note that these studies repeated these items several times in blocks, however. We opted against using repetition in our design, given the evidence that emotional words in particular are susceptible to immediate and delayed repetition priming effects (D. Zhang et al., 2019). In addition, our reported power analysis indicates sufficient sample size even with these lower item counts. Future studies should employ higher more trials. (2) We matched multiple word properties (Table 1), but we did not match part of speech, concreteness, or familiarity. What part of
speech, a swear word is not easily be determined. The majority of swear words can be	nouns (What is this shit?), verbs (He shit himself.), adjectives (That was a shit film), or
virtually any other part of speech. In this study, we compared swears with closed-class
function words to test the claim in the literature that swears act as closed-class words. In
some sense, it is part of the research question, as opposed to being a variable that must be
controlled.

Concreteness of swears is also hard to pin down and likely context dependent (e.g.,
consider shit in This song is my shit and My dog took a shit). For example, in This song is
my shit, shit is very abstract, but in My dog took a shit, shit is concrete and reflects literal
faecal matter. In Kanske and Kotz (2007), concreteness interacted with emotion in that
only concrete, negative words (e.g. bomb) showed the enhanced LPC, but not abstract
negative words (e.g. violence).

With respect to familiarity, we matched the word frequency, which according to some
school of thought (Tanaka-Ishii & Terada, 2011) can be viewed as an approximant of
familiarity. In addition, this is especially so given that our stimuli have a relatively high
frequency, which typically have high familiarity. The familiarity effect matters most for low
frequency items. Finally, in our data, participants responded with over 90% accuracy to the
critical comparison conditions swear, negative, and neutral, suggesting that participants were
familiar with these stimuli. The reduced accuracy for the closed condition might possibly
suggest that familiarity played a role here, but these words were just as frequent as the other
conditions, suggesting that participants have encountered these items on a regular basis.

5. Conclusion

The current study found that swear words have content distinct from function and
from negative words, and that this content is emotional and social. Specifically, swears and
negative words elicited a larger positivity (250-550 ms) than open-class neutral words,
associated with attention due to negative valence. Later on, swears elicited a larger late positivity (550-750 ms) than negative words, associated with social tabooness. Our findings suggest that a two-dimensional model of emotional language processing requires expansion in order to fully capture all emotional language including swears. With these results, we contribute a new data set for probing our understanding of one of the central properties of language: conveying emotion.

References


URL: http://mc.manuscriptcentral.com/pcem Email: PCEM-peerreview@journals.tandf.co.uk


URL: http://mc.manuscriptcentral.com/pcem Email: PCEM-peerreview@journals.tandf.co.uk


Stephens, R., Atkins, J., & Kingston, A. (2009). Swearing as a response to pain: *NeuroReport, 1.* https://doi.org/10.1097/WNR.0b013e32832e64b1


## Tables

Table 1. Mean ratings/values with standard deviations in parentheses of word properties for each condition.

<table>
<thead>
<tr>
<th>Word Properties</th>
<th>Swear</th>
<th>Negative</th>
<th>Open Class</th>
<th>Closed Class</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Valence</td>
<td>2.46 (.92)</td>
<td>2.04 (.65)</td>
<td>5.9 (1.01)</td>
</tr>
<tr>
<td></td>
<td>Arousal</td>
<td>5.57 (.94)</td>
<td>5.44 (.78)</td>
<td>3.81 (.81)</td>
</tr>
<tr>
<td></td>
<td>Tabooneess</td>
<td>6.82 (.84)</td>
<td>5.03 (.90)</td>
<td>2.95 (.37)</td>
</tr>
<tr>
<td></td>
<td>Log Subtlex Word Freq</td>
<td>3.31 (.68)</td>
<td>3.58 (.4)</td>
<td>3.25 (.52)</td>
</tr>
<tr>
<td></td>
<td>Letter Length</td>
<td>4.85 (1.32)</td>
<td>4.85 (.80)</td>
<td>5 (.85)</td>
</tr>
<tr>
<td></td>
<td>Num Morphemes</td>
<td>1.27 (.44)</td>
<td>1.22 (.42)</td>
<td>1.11 (.31)</td>
</tr>
<tr>
<td></td>
<td>Num Phonemes</td>
<td>3.81 (1.04)</td>
<td>4.11 (.87)</td>
<td>4.04 (.78)</td>
</tr>
<tr>
<td></td>
<td>Num Syllables</td>
<td>1.27 (.44)</td>
<td>1.52 (.5)</td>
<td>1.5 (.5)</td>
</tr>
</tbody>
</table>
**Figure 1.** Jittered sequence of one trial of the lexical decision task. Duration of each target depends on letter length.

**Figure 2.** Proportion of correct responses for lexical decision task experimental conditions.

**Figure 3.** Mean reaction time for experimental conditions.

**Figure 4.** Grand averaged ERP waveforms for experimental conditions at the frontal, central, and posterior locations.

**Figure 5.** Scalp topographies at for earlier (250-550) and LPC (550-750) time windows. All voltage differences are significant unless otherwise noted.

**Figure 6.** ERP waveforms and scalp topographies for pseudoword and open-class neutral words.
Sequence of one trial of the lexical decision task.

Figure 1. Jittered sequence of one trial of the lexical decision task. Duration of each target depends on letter length.

338x190mm (96 x 96 DPI)
Figure 2. Proportion of correct responses for lexical decision task experimental conditions.

338x190mm (96 x 96 DPI)
Figure 3. Mean reaction time for experimental conditions.

338x190mm (96 x 96 DPI)
Figure 4. Grand averaged ERP waveforms for experimental conditions at the frontal, central, and posterior locations.

338x190mm (96 x 96 DPI)
Figure 5. Scalp topographies at for earlier (250-550) and LPC (550-750) time windows. All voltage differences are significant unless otherwise noted.

338x190mm (96 x 96 DPI)
Figure 5. Scalp topographies at for earlier (250–550) and LPC (550–750) time windows. All voltage differences are significant unless otherwise noted.

338x190mm (96 x 96 DPI)
Figure 6. ERP waveforms and scalp topographies for pseudoword and open-class neutral words.