Affective biases in English are bi-dimensional

Amy Beth Warriner\textsuperscript{a} & Victor Kuperman\textsuperscript{b}

\textsuperscript{a} Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, ON, Canada
\textsuperscript{b} Department of Linguistics and Language, McMaster University, Hamilton, ON, Canada

Published online: 14 Oct 2014.

To cite this article: Amy Beth Warriner & Victor Kuperman (2014): Affective biases in English are bi-dimensional, Cognition and Emotion, DOI: 10.1080/02699931.2014.968098

To link to this article: http://dx.doi.org/10.1080/02699931.2014.968098
Affective biases in English are bi-dimensional

Amy Beth Warriner1 and Victor Kuperman2

1Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, ON, Canada
2Department of Linguistics and Language, McMaster University, Hamilton, ON, Canada

A long-standing observation about the interface between emotion and language is that positive words are used more frequently than negative ones, leading to the Pollyanna hypothesis which alleges a predominantly optimistic outlook in humans. This paper uses the largest available collection of affective ratings as well as insights from linguistics to revisit the Pollyanna hypothesis as it relates to two dimensions of emotion: valence (pleasantness) and arousal (intensity). We identified systematic patterns in the distribution of words over a bi-dimensional affective space, which (1) run counter to and supersede most prior accounts, and (2) differ drastically between word types (unique, distinct words in the lexicon) and word tokens (number of occurrences of available words in the lexicon). We argue for two factors that shape affect in language and society: a pro-social benevolent communication strategy with its emphasis on useful and dangerous phenomena, and the structure of human subjective perception of affect.

Keywords: Valence; Arousal; Lexicon; Cognitive bias; Subjective experience.

Researchers have long explored the emotional structure of the lexicon and the emotionality of word choices in individuals as the basis for identifying affective characteristics of individual language users and language communities. For example, older people are shown to use more positive, future tense and complex words than younger people (Pennebaker & Stone, 2003), and females use more tag questions and intensifiers (Lakoff, 1975) while males use more judgmental adjectives and directives (Mulac, Braden, & Gibbons, 2001). Word choice can also reveal personality—the frequency of use of negative words increases with neuroticism while that of positive words increases with extraversion (Pennebaker & King, 1999); emotion—anxiety comes with the use of explainers and negations while anger is related to a lack of qualifiers (Weintraub, 1981, 1989); and affective disorders—depressed individuals use more first person singular pronouns (Bucci & Freedman, 1981; Rude, Gortner, & Pennebaker, 2004). The effect of word choices can even reveal what people...
wish to hide in that liars can be detected by their increased use of negative words (Newman, Pennebaker, Berry, & Richards, 2003; Vrij, 2000) among other factors. With increased computing power has come the ability to examine the link between linguistic behaviour, emotional dynamics and cultural values of large social groups through extensive collections of language samples, both written and spoken. Examples include new research fields such as opinion mining and sentiment analysis: large-scale data-mining studies were conducted to measure attitudes towards products and events, through mostly online texts such as websites, forums and twitter feeds (for a review, see Liu, 2012). Another instance of an emerging relevant field is culturomics which analyses digital texts, observing cultural trends and their change over time through distributions of word frequencies (Michel et al., 2011).

The present paper takes, as its point of onset, an observation about language and its cultural and emotional characteristics that informs all research disciplines mentioned above, namely that positive words (those with higher psychological valence) occur more frequently than negative words (those with lower valence). Boucher and Osgood (1969) are credited as being the first to report this observation about the English language: they dubbed the phenomenon the “Pollyanna hypothesis” after the title character of a series of children’s novel known for her invariably optimistic outlook. Multiple further studies (see the review below) have confirmed the Pollyanna or the positivity bias across languages. Importantly, however, the use of the term positivity bias has come to conflate two theoretically and empirically distinct phenomena: (1) positive words are used more frequently than negative words and (2) language has more distinct positive words than distinct negative words (see Kloumann, Danforth, Harris, Bliss, and Dodds (2012), and references in the list). To rephrase the two observations in linguistic terms, the positivity bias may express itself as a correlation of valence with (1) token frequency or the number of instances of a given word observed in a sample of text; (2) type frequency or the number of distinct, unique words (regardless how often each of them is used) in a sample of text; or both (1) and (2). To clarify the difference between the two frequencies with an example, in the sentence “The cat ate the mouse”, the word “the” has a type frequency of 1 and a token frequency of 2. Later, we cast our review of the prior literature in terms of token and type frequency and their link with positivity. We follow the review with a discussion of the empirical and theoretical implications of this distinction.

Several contemporaries of Boucher and Osgood (1969) confirmed their observation of a correlation between the positivity of a word and its token frequency across languages. For example, a year earlier, Zajonc’s (1968) paper on the mere exposure effect noted that affective connotation was related to word frequency. The possibility that positivity affects word type frequency and leads to a prevalence of positive words in language was not considered by Boucher and Osgood (1969); however, just a few years earlier, Johnson, Thomson and Frincke (1960) reported that the ratio of positive to negative words in their sample was 2:1.

In recent years, several researchers have set out to confirm that a positivity bias continues to exist now, 40 years later. Augustine, Mehl, and Larsen (2011) used words that had been rated for both valence (negative vs. positive) and arousal (calm vs. excited) in the Affective Norms for English Words (ANEW) study (Bradley & Lang, 1999). They correlated these ratings with both a small set of frequency norms from Kucera and Francis (1967; 1.014 million word token corpus) and a much larger set of frequency norms from the Hyperspace Analogue to Language (HAL; Lund & Burgess, 1996; 160 million word token corpus).

Regardless of which corpus was used, valence was strongly related to token frequency but arousal was
not. Thus, despite the fact that the percentages of distinct positive and negative words in the sample were nearly equal (i.e., the type frequencies of positive and negative words were similar), on average positive words appeared far more often (i.e., had more tokens) in the two corpora than negative words. Augustine et al. (2011) extended past findings by showing that the correlation between valence and token frequency was significant across different parts of speech (nouns, verbs and adjectives), and within a sample of spoken speech. It is worth noting that the results of Augustine et al.’s study, as well as any study using the ANEW ratings of emotionality, may be confounded by the fact that lexical stimuli of the ANEW study were specifically chosen to equally represent the entire affective space, including its extremes (Bradley & Lang, 1999), rather than represent the distributions of affective words as observed in natural language. Data-sets that select stimuli solely based on their frequency of occurrence (e.g., Kloumann et al., 2012 and Warriner, Kuperman & Brysbaert, 2013) do not introduce this distributional bias.

Rozin, Berman, and Royzman (2010) took a different approach and examined a small set of adjectives and nouns denoting emotional states (disgust, sympathy) by interviewing native speakers of 20 different languages to identify patterns in their use. They determined that there was a distinct advantage for positive words in all studied languages: positive words are more typically unmarked (“not bad” is more common than “not good”) and more likely to be reversed to form their opposite (“unhappy” is more common than “unsad”). Negated positive adjectives (“not pretty”) are viewed as negative while negated negative adjectives (“not ugly”) are viewed as neutral. Also when paired, positive adjectives are typically mentioned first (e.g., “good and bad”, “pros and contras”). This advantage associated with positive words forms a corollary to the fact that they are observed more frequently; there are more tokens of each type. Notably, Rozin et al. (2010) also report a word-type negative bias: negative emotion labels are more diversified and lexicalised in a larger number of languages as compared to their positive antonyms. That the specific lexical space of words that label emotions (e.g., pleasure and disgust) contains more negative than positive types of emotion labels is further confirmed in corpus studies and in free-listing experiments (Schrauf & Sanchez, 2004; Semin & Fiedler, 1992; Russell, 1991). The reported negativity bias in emotion labels appears to run counter to the widely reported positivity bias in the lexicon at large and requires an explanation.

A correlation between positivity and word token frequency was also replicated in written English by Unkelbach et al. (2010) and in Italian adjectives by Suittner and Maas (2008). Garcia, Garas, and Schweitzer (2012) similarly demonstrated a strong positive relationship between valence and token frequency in English, German and Spanish words. Despite this accumulation of evidence, one significant limitation in all these studies has been the small set of emotional norms upon which the researchers have been able to draw. The largest of these studies are those that used the 1034 English word norms in ANEW (Bradley & Lang, 1999).

Addressing this limitation, Kloumann et al. (2012) considered the top 5000 most frequently used words from each of four different corpora, collecting valence ratings for the resulting list of 10,222 unique word types. They confirmed a positivity bias in all four corpora—Twitter, Google Books, New York Times and music lyrics—at the word type level. The mean of the positivity ratings of the top 5000 word types in each of these data-sets was greater than the valence scale midpoint, suggesting that more word types are positive than negative. However, being at odds with the majority of previously published work, Kloumann et al. disconfirmed the well-established correlation of positivity with word token frequency, as they only observed a weak relationship between word valence and token frequency in their Twitter and music lyrics sources and an even weaker correlation in Google Books and the New York Times.

---

2 We are indebted to an anonymous reviewer for raising this point.
They concluded that the two dimensions of language use reflected by type and token frequencies should be considered independent, and that the positive outlook reflects itself in a richer gamut of positive rather than negative experiences, and not in the fact that people tend to mention their positive experiences more often than negative ones. To sum up, what was strong cross-linguistic support for a relationship between emotionality and token frequency (we use a word more frequently if it is more positive) has become contested in Kloumann et al. (2012) and the possibility of a type-based positivity bias (more words are positive) received its first strong evidence.

As argued earlier, the distinction between type and token frequencies is theoretically important, as token and type frequencies of occurrence reflect different mechanisms in language use (see examples in Bybee, 2010). If a language’s word-stock is metaphorically construed as a toolkit, with each word as a tool developed to satisfy a specific communicative need, then a higher type frequency of positive words means that there is a need to have more diverse tools to express phenomena related to positive experiences. A higher token frequency of positive words means, however, that the tools existing for relatively positively phenomena are more in demand than those for relatively negative phenomena. Logically, the number of unique tools and how often some of the tools are used might not be related, and so the correlations of positivity with type and token frequencies of words need to be considered and interpreted independently in order to characterise the interplay of language and emotion.

Another reason for treating type and token frequencies as empirically and theoretically distinct concepts goes beyond the need for factual accuracy in characterising affective language use. Rather it derives from the fact that they give rise to different causes for the positivity bias. The preference for a larger number of positive words (i.e., the type-based bias) in language may reflect a broader diversity of positive than negative phenomena in cumulative human experience (Gable, Reiss, & Elliot, 2000; Rozin et al., 2010). Alternatively, the prevalence of positive word types may indicate a stronger communicative need to express fine-grained semantic aspects of positive phenomena than negative ones: as the semantic growth model by Steyvers and Tenenbaum (2005) argues, semantic differentiation of word meanings in language primarily affects words that are in heavy use and is achieved via creation of new word types with more specific meanings. Finally, the type-based bias may be informed by both the emotional structure of human experience and the communicative needs of language users. Interestingly, while Boucher and Osgood (1969) and much subsequent work entertained the prevalence of positive phenomena as an explanation for the positive outlook, very few of these studies have actually explored the type frequency of positivity.

Conversely, the token-based positivity bias, i.e., a more frequent use of a typical positive word, does not reflect the spectrum of emotional possibilities but rather suggests a preferred selection of relatively positive meanings from the available spectrum. This optimistic bias does not only characterise human communication, where it is described as pro-social benevolent behaviour (Augustine et al., 2011), but is also observed in the human tendency to make psychological and economic predictions that overestimate reality: e.g., the likelihood of divorce, employment prospects and children’s success are often anticipated to be more favourable than is warranted by statistics (see Sharot, 2011 for biological and evolutionary underpinnings of this bias). Intriguingly, while Kloumann et al. (2012) resort to pro-social behaviour as an explanation of the distributional patterns in their data, they only find a weak token-based positivity bias, i.e., the correlation that is an index of pro-social behaviour.

Apart from revealing the emotional structure of language, the existence of the bias appears to have behavioural consequences. For instance, negative stimuli have an advantage when it comes to grabbing attention, being remembered and appearing more potent (for reviews, see Rozin & Royzman, 2001; Baumeister, Bratslavsky, Finkenauer, & Vohs 2001). However, positive information tends to be classified, evaluated and responded to faster (Bargh, Chaiken, Govender, & Pratto, 1992; Unkelbach et al., 2010,
The linguistic positivity bias with its possible and separable links between psychological valence and word type and token frequencies suggests a unifying explanation. Negative word types are rarer and hence marked (Clark & Clark, 1977) which leads to improved attention and memory performance for negative words. Positive words have a higher token frequency and thus carry less information (Garcia, et al., 2012), are more densely clustered in the lexicon and are therefore privileged by faster and greater spreading activation (Unkelbach, Fiedler, Bayer, Stegmuller, & Danner, 2008; Unkelbach, 2012). A potential explanatory power of the positivity bias requires understanding of the nature of this systematic relationship between language and emotion, and particularly the aspects of linguistic use that it affects: lexical diversity (roughly corresponding to type frequency), lexical entrenchment (reflected in word token frequency) or both. Our first goal is to utilise a recently collected, large set of emotional norms for 13,915 words (Warriner et al., 2013) to definitively characterise a positivity bias in the English language and determine its relationship to word type and token frequencies, and its implications for the emotional tenor of society. A type bias will be evident if more than 50% of words in the corpora studied are positive and a token bias will be evident if there is a reliable positive correlation between valence and word frequency in those same corpora.

Emotion, however, is more than negativity vs. positivity. An influential view on the structure of emotion holds that emotion is dimensional and that the two primary dimensions are those of valence (a negative vs. positive emotional state) and arousal (a calm vs. excited state; for reviews, see Barrett and Russell (1999); Fontaine, Sherer, Roesch, and Ellsworth (2007); and Power (2006). Furthermore, while proposals vary widely as to the nature of the relationship between these dimensions, arousal is often (partly or wholly) associated with the intensity of pleasure or displeasure that one experiences in response to a stimulus (for a review of theories, see Kuppens, Tuerlinckx, Russell, & Barrett, 2013). Most findings in the realm of subjective experience—i.e., self-reports of affect elicited, among others, in the form of ratings—support a characteristic “boomerang”-shaped functional relationship between valence and arousal (Bradley & Lang, 2007, 2009). Namely, stimuli subjectively perceived as very negative or very positive tend to come with a high level of arousal, while mildly positive or negative stimuli are perceived as relatively calm. The boomerang- or the U-shaped relationship is found in aggregate ratings to a broad range of stimuli types, including emotionally laden pictures, affective experience in daily life, current and remembered affective experiences (Kuppens et al., 2013) and, importantly, words (cf. Bradley & Lang, 1999; Redondo, Fraga, Padrón, & Comesaña, 2007; Soares, Comesaña, Pinheiro, Simões, & Frade, 2012; Warriner et al., 2013). While opposing accounts exist, the psychological basis of this relationship is thought to be that subjective valence of a stimulus engages one of two motivational subsystems: an appetitive/positive one geared towards attaining objects beneficial for survival (e.g., sustenance, nurturance and caregiving) and an averse/ negative one associated with response to threat and danger (Bradley, 2000; Bradley & Lang, 2000). Arousal then is a metric of how strongly these systems are activated and how much effort needs to be mobilised to respond to environmental demands (for early proposals, see Duffy, 1951; Kahneman, 1973). The effort is maximal in extremely positive or negative cases and minimal when stimuli are neutral and the activation level of motivational systems is low (Kuppens et al., 2013).

The well-established systematic relationship between valence and arousal in indices that aggregate subjective experience of affect across multiple individuals naturally raises the following questions: Does language have a larger number of calm than exciting words and are exciting words used more frequently than their calmer counterparts? Do distributional patterns of positivity vary by the level of arousal? Are bi-dimensional patterns different for word types and tokens? Our second goal then is to explore—using this same large set of lexical affective norms—the possibility of a type-based and a token-based bias with regard
to the arousal of words in English, and ultimately a compound bi-dimensional distributional bias influenced both by valence and arousal.

METHOD

Emotional ratings

We used Warriner et al.’s (2013) collection of valence and arousal ratings for 13,915 lemmas, or vocabulary word forms. For instance, sing is a lemma for such inflected word forms as sing, sang, sung and singing; thus, a lemma merges grammatical variants of the word, which are not expected to vary emotionally. These ratings were collected using the crowd-sourcing online Amazon Mechanical Turk platform (Schnoebelen & Kuperman, 2010) and validated via correlations with previously collected ratings. Participants were drawn from native English-speaking residents of the USA. Each was assigned to rate a sample of words on a 9-point scale ranging either from sad (1) to happy (9) or from calm (1) to excited (9). At least 18 participants rated the majority of words and overall means per word were calculated: one for valence and one for arousal. The words were selected on the basis of their familiarity to a typical English speaker. They were drawn from the list of 30,000 words that were indicated as known by at least 70% or more participants in the study by Kuperman, Stadthagen-Gonzalez, and Brysbaert (2012). All lemmas represented content words (nouns, verbs and adjectives) rather than function words (“the”, “this” and “from”). For overall analyses in the current paper, we used the 13,763 word types from Warriner et al. (2013) for which there was frequency information in the 51 million-token SUBTLEX-US corpus based on subtitles to US films and media (Brysbaert & New, 2009); altogether, the word types chosen accounted for 14.5 million tokens.

Other corpora

We used multiple additional corpora in order to explore genre, regional and age variability, and thus validate the generalisability of our findings. Type and token frequency measures were drawn from these corpora and correlated with the emotional ratings from Warriner et al. (2013). We analysed the following corpora: the Corpus of Contemporary American English (COCA; Davies, 2009)—a 450 million token balanced corpus including spoken, fiction, magazines, newspapers and academic texts from 1990 to 2012; the British National Corpus (BNC; Oxford University Computing Services, 2007)—a 100 million token collection of the UK-based written and spoken language from the 1990s; the Touchstone Applied Science Associates Corpus (TASA; Zeno, Ivens, Millard, & Duvvuri, 1995)—12 million tokens from textbooks, literature and novels separated by cumulative grade level from grade 3 to college; and the HAL (Lund & Burgess, 1996) frequency norms—roughly 160 million token gathered from Usenet groups in 1995. The number of words that overlapped between each corpus and the rating study are listed in Column N of Table 1.

RESULTS AND DISCUSSION

In keeping with prior research, we first report the results of considering the valence and arousal biases independently, as linear predictors of word frequency expressed in types and tokens. We then demonstrate a role of non-linearity in characterising these effects and finally proceed to analyses of the compound valence-and-arousal bias.

Independent positivity and arousal biases: Type frequency

In our overall data-set (words and frequency drawn from SUBTLEX-US), the distribution of valence ratings showed a longer left tail and a slight negative skew (−0.29). Both the mean (5.06) and the median (5.20) were above 5, which is the mid-point of the scale used. Overall, 55.6% of the words were positive, i.e., had valence ratings above the scale’s mid-point, confirming the widely reported positivity type bias, see the histogram in Figure 1. We also confirmed that the positivity bias holds when a subset of more extreme valence ratings is
Table 1. Summary of type and token frequencies for various datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Percent of positive word types</th>
<th>Median valence</th>
<th>Skewness valence</th>
<th>Correlation of valence with token frequency</th>
<th>Median arousal</th>
<th>Skewness arousal</th>
<th>Correlation of arousal with token frequency</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall (SUBTLEX)</td>
<td>55.6</td>
<td>5.20</td>
<td>−0.29</td>
<td>0.180</td>
<td>4.11</td>
<td>0.51</td>
<td>0.039</td>
<td>13,763</td>
</tr>
<tr>
<td>Overall – Nouns</td>
<td>58.5</td>
<td>5.25</td>
<td>−0.42</td>
<td>0.172</td>
<td>4.05</td>
<td>0.57</td>
<td>0.044</td>
<td>8,202</td>
</tr>
<tr>
<td>Overall – Verbs</td>
<td>52.3</td>
<td>5.00</td>
<td>−0.01</td>
<td>0.236</td>
<td>4.27</td>
<td>0.35</td>
<td>−0.007</td>
<td>1,753</td>
</tr>
<tr>
<td>Overall – Adjs</td>
<td>49.1</td>
<td>5.10</td>
<td>−0.21</td>
<td>0.149</td>
<td>4.15</td>
<td>0.47</td>
<td>0.086</td>
<td>3,129</td>
</tr>
<tr>
<td>COCA</td>
<td>55.6</td>
<td>5.20</td>
<td>−0.29</td>
<td>0.247</td>
<td>4.11</td>
<td>0.51</td>
<td>−0.053</td>
<td>13,762</td>
</tr>
<tr>
<td>COCA – Fiction</td>
<td>71.5</td>
<td>5.67</td>
<td>−0.59</td>
<td>0.222</td>
<td>4.05</td>
<td>0.55</td>
<td>−0.109</td>
<td>3,158</td>
</tr>
<tr>
<td>COCA – Magazines</td>
<td>76.7</td>
<td>5.72</td>
<td>−0.72</td>
<td>0.167</td>
<td>4.00</td>
<td>0.61</td>
<td>−0.069</td>
<td>3,472</td>
</tr>
<tr>
<td>COCA – Spoken</td>
<td>69.6</td>
<td>5.59</td>
<td>−0.64</td>
<td>0.126</td>
<td>4.14</td>
<td>0.53</td>
<td>−0.043</td>
<td>3,424</td>
</tr>
<tr>
<td>COCA – Newspapers</td>
<td>73.3</td>
<td>5.68</td>
<td>−0.70</td>
<td>0.135</td>
<td>4.05</td>
<td>0.57</td>
<td>−0.046</td>
<td>3,383</td>
</tr>
<tr>
<td>COCA – Academic</td>
<td>74.4</td>
<td>5.61</td>
<td>−0.72</td>
<td>0.132</td>
<td>4.00</td>
<td>0.61</td>
<td>−0.032</td>
<td>3,302</td>
</tr>
<tr>
<td>BNC</td>
<td>56.4</td>
<td>5.21</td>
<td>−0.29</td>
<td>0.242</td>
<td>4.05</td>
<td>0.54</td>
<td>−0.068</td>
<td>7,823</td>
</tr>
<tr>
<td>TASA – Grade 3</td>
<td>60.3</td>
<td>5.32</td>
<td>−0.36</td>
<td>0.209</td>
<td>4.05</td>
<td>0.50</td>
<td>−0.093</td>
<td>8,301</td>
</tr>
<tr>
<td>TASA – Grade 6</td>
<td>57.9</td>
<td>5.25</td>
<td>−0.32</td>
<td>0.231</td>
<td>4.09</td>
<td>0.51</td>
<td>−0.101</td>
<td>11,185</td>
</tr>
<tr>
<td>TASA – Grade 9</td>
<td>57.3</td>
<td>5.24</td>
<td>−0.31</td>
<td>0.233</td>
<td>4.09</td>
<td>0.51</td>
<td>−0.098</td>
<td>11,693</td>
</tr>
<tr>
<td>TASA – Grade 12</td>
<td>56.8</td>
<td>5.23</td>
<td>−0.30</td>
<td>0.241</td>
<td>4.10</td>
<td>0.51</td>
<td>−0.105</td>
<td>12,135</td>
</tr>
<tr>
<td>TASA – College</td>
<td>56.4</td>
<td>5.21</td>
<td>−0.29</td>
<td>0.235</td>
<td>4.10</td>
<td>0.51</td>
<td>−0.104</td>
<td>12,344</td>
</tr>
<tr>
<td>HAL</td>
<td>55.9</td>
<td>5.20</td>
<td>−0.29</td>
<td>0.209</td>
<td>4.10</td>
<td>0.51</td>
<td>−0.037</td>
<td>12,952</td>
</tr>
</tbody>
</table>

Columns 2–4 reports the percent of words above the midpoint of the positivity scale, as well as the median value and skewness of valence; columns 6–7 report the median value and skewness for arousal. These measures represent type frequencies. Columns 5 and 8 report Spearman’s correlations between each of the two emotional dimensions—valence and arousal—from the Warriner et al. (2013) study and the word token frequency of that particular corpus. Column 9 indicates how many word types in that corpus had emotional ratings available in Warriner et al. (2013).
considered, rather than the entire data-set with numerous valence-neutral items: the number of words with valence above 6 was 10% larger than the number of words with valence below 4 (3106 vs. 2869).

By contrast, the distribution of arousal ratings showed a strong positive skewness (0.51), indicating that the majority of the probability mass is concentrated on the lower end of the 1–9 arousal scale, see Figure 1. The mean and median ratings of arousal were 4.21 and 4.11, respectively. These patterns point to the preference towards calmer rather than more exciting words. We do not draw a comparison of arousal ratings against this scale’s mid-point, since in the case of a unipolar psychological dimension like arousal, this point is not informative.

Skewness in similar directions was found in the other corpora examined (see Table 1 for a summary), which is not surprising given how many words overlapped across corpora. UK English (BNC) and US English (HAL and COCA) showed nearly equivalent type biases (i.e., median and level of skewness almost identical) for both valence and arousal. To test variability across genres, we further created subsets of 4000 words representing most frequent nouns, verbs and adjectives for each of the five genres of COCA. Arousal medians and skewness in each of the genres were nearly equivalent to that of COCA as a whole. Together these similarities suggest the robustness of the positivity and calmness type-based biases across dialects and genres. The biases were also robust developmentally, i.e., across grade levels of the TASA corpus: nearly identical skewness was found for both arousal and valence with a small numerical decrease in magnitude for valence as grade level increases.

We also examined how the type bias was manifested within the three parts of speech we selected from SUBTLEX-US. For nouns (N = 8802), there was a significant positive type bias with 58.5% being positive ($\chi^2 = 254.38, p < .0001$). For verbs (N = 1753), the positive bias was only marginally significant with 52.3% being positive ($\chi^2 = 3.71, p = .054$), while for adjectives (N = 3,129), there was no type bias in any direction with only 49.1% being positive ($\chi^2 = 1.01, p = .314$). For all parts of speech, distributions of arousal were positively skewed, indicating a strong bias towards calmer word types. This skew was strongest in
nouns (0.57) followed by adjectives (0.47) and then verbs (0.35).

To sum up, the percentage of positive words ranged across corpora from 55.6 to 76.7. Thus, we confirm, across corpora, the previously reported tendency (Kloumann et al., 2012) for English to have a larger diversity of word types with positive rather than negative connotations. We further determined that this type bias is more nuanced than reported earlier, in that it is present mostly within nouns, and somewhat within verbs. There is no positivity type bias for adjectives. Given the prototypical semantic roles of parts of speech in a language, this suggests that the positivity bias is largely driven by a stronger representation of positive rather than negative objects (nouns) in the word stock. There is however a roughly equivalent representation of positive and negative actions and states (verbs), and object qualities (adjectives). Some of our findings run counter to results reported by Kloumann et al.’s (2012), which offers a similar statistical power and diversity of written genres of English. We address methodological reasons for these discrepancies in the Appendix.

We report, for the first time, a strong and consistent tendency for English to have a greater variety of words for expressing calm things across all parts of speech, and to offer a larger toolkit for expressing calm-evoking objects, actions and traits.

Independent positivity and arousal biases: Token frequency

In Warriner et al.’s (2013) data-set, we found the Spearman’s correlation between valence and log SUBTLEX-US frequency to be $\rho = 0.180$, i.e., stronger than the correlations reported in Kloumann et al. (2012), see Table 1 and Figure 1. Together, the fact that there are more positive words (type bias) and those positive word types occur more often (token bias) leads to a large prevalence of positive words in general. There are 4.4 times more words with valence above the mid-point than those below the mid-point, and 1.7 times more words with valence above 6 than those with valence below 4. For completeness, we note a slight increase in token frequency in very negative words, along with a stronger and steeper increase in token frequency of relatively positive words, see Figure 1. Our examination of other corpora showed similar patterns of correlations (see Table 1 for a summary). The correlation between valence and token frequency ranged from $\rho = 0.126$ to $\rho = 0.247$.

We further observed a negligible Spearman’s correlation between arousal and log SUBTLEX-US token frequency ($\rho = 0.039$). Similarly, correlations between arousal and token frequency across corpora were weak, fluctuated in polarity and ranged between $\rho = -0.109$ and $\rho = 0.034$. This apparent null effect was in line with the finding of Augustine et al. (2011). However, a closer look at the distribution of word tokens over the range of arousal revealed a quadratic functional relationship, with a steep increase in token frequency in very calm and very exciting words, see Figure 1. A similar pattern was confirmed across corpora with the help of regression models: an adjusted $R^2$ for a model fitted to token frequency in SUBTLEX-US with a linear function of arousal as a predictor is 0.1%, while a model with a linear and a quadratic term for arousal yielded an adjusted $R^2$ of 0.6%. Similar significant ($p < .05$) advantages of the quadratic over a linear functional form of the arousal effect were observed in other goodness-of-fit measures (AIC and BIC). Thus, given its symmetrical nature, the zero-order correlation yields a value that is close to zero: yet, this apparent null effect is merely an artefact of imposing linearity on a non-linear relationship between variables (see Kuppens et al., 2013 for elaboration of this methodological point).

To summarise, we observed consistent positive correlations between word token frequency and its emotional valence. The relationship between word token frequency and word arousal was also systematic and had a quadratic U shape.

Stability of independent affective biases across frequency ranges

It is a logical possibility that the type- and token-based biases towards positive and calm words are
specific to certain frequency ranges and would not be observed if very uncommon words were taken out of consideration. We ruled out this possibility by supplementing the analyses of the entire (100%) word list of Warriner et al. (2013) with the analyses of words in the top 90%, 80%, 70% … 10% of the data-set’s token frequency range. For each subset, Table 2 reports skewness coefficients and medians for the type-based distributions of valence and arousal, as well as the percent of positive words, and Spearman’s correlation coefficients for valence and log token frequency. Correlations between arousal and log token frequencies are not reported, as they offer a poor fit to non-linear relationship between the two variables.

Table 2 confirms observations made on the entire data-set of Warriner et al. (2013). The number of positive word types (valence above 5) steadily increases as we move to portions of the data-set containing higher frequency words, and so does the median and absolute value of the skewness of the valence distribution over types. Taken together, these indices suggest that on average word types become more positive (and the left tail of their distribution becomes longer) in the higher frequency ranges. The token-based positivity bias remains stable over the entire data-set, with only a minor non-significant decrease in the magnitude of the correlation between word valence and log frequency in the top decile of the frequency distribution. That is, the token-based positivity bias is not driven by exceedingly rare words or even words in the mid-range of frequency. Finally, a strong type-based bias towards calmer words is confirmed by the virtually invariant estimates of the median and skewness of arousal rating.

### Compound affective bias: Types

As discussed in the Introduction, valence covaries with arousal in a number of tasks that elicit self-reports of subjective experience, including affective ratings to words. The relationship forms a characteristic boomerang functional curve, with higher arousal accompanying extreme values of valence. It is plausible that the observed biases towards positivity and calmness only partially reflect an overarching bias that spans over both affective dimensions. To explore the distribution of unique words over a bi-dimensional affective space, we calculated the number of word types for each of the 100 bins formed by the crossing of deciles

<table>
<thead>
<tr>
<th>Percent of dataset</th>
<th>Percent of positive word types</th>
<th>Median valence</th>
<th>Skewness valence</th>
<th>Correlation of valence with token frequency</th>
<th>Median arousal</th>
<th>Skewness arousal</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.56</td>
<td>5.20</td>
<td>−0.29</td>
<td>0.18</td>
<td>4.11</td>
<td>0.51</td>
<td>13,763</td>
</tr>
<tr>
<td>90</td>
<td>0.57</td>
<td>5.23</td>
<td>−0.32</td>
<td>0.18</td>
<td>4.14</td>
<td>0.50</td>
<td>12,272</td>
</tr>
<tr>
<td>80</td>
<td>0.58</td>
<td>5.25</td>
<td>−0.33</td>
<td>0.18</td>
<td>4.14</td>
<td>0.50</td>
<td>10,994</td>
</tr>
<tr>
<td>70</td>
<td>0.59</td>
<td>5.29</td>
<td>−0.36</td>
<td>0.18</td>
<td>4.14</td>
<td>0.49</td>
<td>9,563</td>
</tr>
<tr>
<td>60</td>
<td>0.60</td>
<td>5.32</td>
<td>−0.38</td>
<td>0.19</td>
<td>4.14</td>
<td>0.49</td>
<td>8,255</td>
</tr>
<tr>
<td>50</td>
<td>0.62</td>
<td>5.37</td>
<td>−0.43</td>
<td>0.18</td>
<td>4.14</td>
<td>0.49</td>
<td>6,873</td>
</tr>
<tr>
<td>40</td>
<td>0.64</td>
<td>5.44</td>
<td>−0.49</td>
<td>0.16</td>
<td>4.13</td>
<td>0.49</td>
<td>5,497</td>
</tr>
<tr>
<td>30</td>
<td>0.67</td>
<td>5.52</td>
<td>−0.54</td>
<td>0.16</td>
<td>4.13</td>
<td>0.49</td>
<td>4,127</td>
</tr>
<tr>
<td>20</td>
<td>0.69</td>
<td>5.61</td>
<td>−0.61</td>
<td>0.16</td>
<td>4.14</td>
<td>0.51</td>
<td>2,753</td>
</tr>
<tr>
<td>10</td>
<td>0.74</td>
<td>5.78</td>
<td>−0.74</td>
<td>0.11</td>
<td>4.15</td>
<td>0.49</td>
<td>1,374</td>
</tr>
</tbody>
</table>

Note: Columns 2–4 reports the percent of words above the midpoint of the positivity scale, as well as the median value and skewness of valence; columns 6–7 report the median value and skewness for arousal. These measures represent type frequencies. Columns 5 and 8 report Spearman’s correlations between each of the two emotional dimensions—valence and arousal—from the Warriner et al. (2013) study and the word token frequency of that particular corpus. Column 9 indicates how many words types are in each bin. Column N indicates how many words are in each bin.
of arousal and valence. Figure 2 is a heatmap of residuals of the chi-squared test associated with each bin.

Resulting patterns reveal that the distribution of word types over valence ratings is strongly modulated by how arousing those words are. Regions of the affective space that accumulate most of the English word-stock are extremely arousing and valenced (positive or negative) words. There is also a large lexical space of calmer words that are relatively neutral in their pleasantness. As such, the word-type distribution in Figure 2 faithfully replicates the boomerang functional curve observed in Bradley and Lang (1999) and confirmed since in multiple studies, including the present Warriner et al.'s set of affective ratings. As more valenced words tend to be more arousing due to a stronger activation of motivational systems, it is not surprising that arousing valenced word types are more prevalent in the language than calm valenced words. Similarly, the tendency for neutrally valenced words to only weakly engage
motivational systems and thus elicit lower levels of arousal translates into a larger number of calm valence-neutral words than exciting valence-neutral words. The visual patterns in Figure 2 are corroborated by the prediction of a generalised additive model with word type frequency as a dependent variable and the tensor product of valence and arousal as a predictor. The model (not shown) fitted a complex hyperbolic surface to the observed data in the three-dimensional space with arousal, valence and word type frequency as axes and confirmed the statistical significance ($p < .0001$) of a surface that reaches maxima of type frequency in extremely arousing, extremely valenced region of the affective space, as well as in the region characterised by low arousal and relatively neutral valence.

The replication of the boomerang shape in the distribution of word types (Figure 2) highlights the inadequacy of considering an inherently bi-dimensional affective phenomenon as a uni-dimensional one. Columns 1–3 in Table 3 illustrate this point by showing how (1) a well-established positivity bias in word types holds true for 60% of words with lower arousal (there are significantly more words with valence >5 as indicated by the proportion test), (2) the bias is not observed in either direction in words falling into deciles 7 and 8 of arousal, and (3) the bias reverses in highly arousing words leading to an advantage to negative word types in this corner of the affective space.

### Compound affective bias: Tokens

Equally nuanced is the token-based positivity bias. Figure 3 reports average log-transformed token frequency for the 100 bins formed by crossing deciles of valence and arousal.

Figure 3 points to highly positive words (deciles 9 and 10) across all arousal levels as the area of the lexicon that is most actively used in communication. This area is complemented by high token frequency words associated with extremely low valence and extremely high arousal, i.e., danger words (see selection (A) in Figure 2). This spike in words reflecting danger may partly explain the paradoxical discrepancy between an overall positivity bias in word types of the entire language, and a negativity bias observed in words serving as emotion labels (Russel, 1991; Schrauf & Sanchez, 2004; Semin & Fiedler, 1992). If a larger number of emotion labels reflect higher arousal states, the preponderance of negative labels is compatible with our findings.
As discussed above, the bi-dimensional patterns further point to a limitation of any study that approaches characterisation of affect in language through the lens of any one dimension. Columns 4 and 7 in Table 3 demonstrate a drastic change in the correlation between valence and token frequency estimated per arousal decile (invariably positive but substantially weakening in higher-arousal words) as well as the correlation between arousal and token frequency (negative and much stronger in negative words than in positive ones).

**Figure 3.** A heat map showing average log frequency per 100 bins formed by crossing valence with arousal deciles. The average log frequency of the words in each bin was calculated and plotted according to the colour key provided. Regions of particular interest, complementary to those in Figure 2, were identified with letters. The table below provides examples of words pulled from each region.

<table>
<thead>
<tr>
<th>(G)</th>
<th>(H)</th>
<th>(I)</th>
<th>(J)</th>
</tr>
</thead>
<tbody>
<tr>
<td>belch</td>
<td>absent</td>
<td>commodity</td>
<td>buddy</td>
</tr>
<tr>
<td>chili</td>
<td>collagen</td>
<td>duet</td>
<td>charm</td>
</tr>
<tr>
<td>drivel</td>
<td>diet</td>
<td>gown</td>
<td>deserve</td>
</tr>
<tr>
<td>eyewitness</td>
<td>electoral</td>
<td>informative</td>
<td>educate</td>
</tr>
<tr>
<td>firewall</td>
<td>flimsy</td>
<td>ketchup</td>
<td>luck</td>
</tr>
<tr>
<td>incorrigible</td>
<td>gutter</td>
<td>lumberjack</td>
<td>nacho</td>
</tr>
<tr>
<td>midterm</td>
<td>overcast</td>
<td>mammal</td>
<td>recipe</td>
</tr>
<tr>
<td>offense</td>
<td>renal</td>
<td>person</td>
<td>shiny</td>
</tr>
<tr>
<td>socialism</td>
<td>unlisted</td>
<td>route</td>
<td>waterfall</td>
</tr>
<tr>
<td>wrinkle</td>
<td>vertigo</td>
<td>waterproof</td>
<td>yummy</td>
</tr>
</tbody>
</table>

**GENERAL DISCUSSION**

The present study explores two novel aspects of the relationship between emotion and language using the largest available data-set of affective ratings (Warriner et al., 2013). First, we revisit the widely reported positivity bias or Pollyanna hypothesis—the prevalence of positive over negative words in language—using the linguistically and psychologically motivated distinction between token frequency (the number of occurrences of a
specific word in a language sample) and type frequency (the number of different words in a language sample; Clark & Clark, 1977; Semin & Fiedler, 1992). Types can be thought of as different tools available for a certain purpose (in this case, for expressing meanings), while tokens can be thought of as how often a given tool is used (how often a certain meaning needs to be expressed). The number of tools and how often specific tools are used are independent metrics of language use, and so identifying a bias in one does not presuppose nor exclude a bias in the other. With these two ideas firmly separated, we report distributional patterns of affect over the word-stock of English using a broad multi-genre selection of corpora and a new comprehensive set of affective ratings (Warriner et al., 2013). The second novel aspect of our study is the expansion of the traditional uni-dimensional view of affective bias in language as a function of valence to a bi-dimensional view, in which both valence and arousal play a role. In the next two sections, we summarise our findings for distributional biases observed in affective dimensions of valence and arousal considered independently, and then we discuss the bi-dimensional affective bias as an overarching pattern.

Independent biases: A consideration of the word’s valence as a sole independent affective dimension—in line with prior literature—reveals that there are a larger number of positive words (above the mid-scale of valence) in the English language, and overall positive words tend to be in use more often (Figure 1). Thus, we confirm both the type-based positivity bias (in agreement with Kloumann et al., 2012) and the token-based positivity bias (contra Kloumann et al., 2012, and in agreement with Boucher & Osgood, 1969 and subsequent cross-linguistic reports): for treatment of discrepancy with Kloumann et al.’s findings see Appendix. The type-based trend towards positivity is further qualified by the fact that the share of positive lexical items significantly exceeds 50% only in English nouns (58.2%), while positive and negative verbs and adjectives tend to be equal in number. If we consider the type-based bias as an indication of the variety of phenomena characterising human experience, the prevalence of positive phenomena is due to words denoting objects (nouns), but not states, motions or qualities (verbs or adjectives). Regardless of the part of speech, English speakers draw upon the positive words more often than the negative ones: while there is a spike in token frequency in very negative words, very happy words are still in more frequent use. An anonymous reviewer pointed out that the positivity bias may be a scaling effect, rather than a reflection of communicative strategies. Negative words may stretch further from a hypothetical valence mid-point than positive words do. For example, while rated similarly, the word “abuse” would intuitively seem to be more extremely negative than “affection” is extremely positive. As a result, people may be breaking up the scale at a point above neutral resulting in neutral words being rated more positively than they actually are. While this is a valid possibility, it is not testable within the constraints of this study. Even though seemingly neutral words like “bathroom” and “lizard” are rated slightly positively, this may be due to the known tendency for people to acquiesce in rating studies (Weijters, Cabooter, & Schillewaert, 2010). Warriner et al. (2013) did not include any reversed scales to counter this effect. We leave the exploration of this possibility to future research, while noting that it is unlikely to underlie the bi-dimensional affective biases that we discuss in this article.

We found, for the first time, a strong type-based bias towards calm words in the English language, observed in all corpora and for all parts of speech (see Figure 1). As such, English speakers have far more calm words available to them than arousing ones. A correlational analysis suggested the absence of a noticeable token frequency bias associated with arousal, in line with Augustine et al. (2011). This, however, only reflects an inability of linear zero-order correlations to approximate non-linear curves (Kuppens et al., 2013). In fact, there is a symmetrical U shape relationship between token frequency and arousal, with very calm and very exciting words being similarly frequent.
Bi-dimensional affective bias: Word types and tokens

Importantly, distributional biases along independent dimensions of affect are superseded by our identification of systematic patterns in the distribution of words over a bi-dimensional affective space formed by valence and arousal. The distribution of word types over valence and arousal axes follows a characteristic boomerang shape, with arousal increasing with extremity of valence (positive or negative). Most word types concentrated around the following regions of affect: high-valence high-arousal (exhilaration, sexual gratification), low-valence high-arousal (danger, threat) and mid-valence low-arousal (emotionally unmarked phenomena; see Figure 2). Regions that are under-represented in word types are those of low-valence low-arousal (depression, shame), high-valence low-arousal (serenity, comfort) and mid-valence high-arousal (see selection (B) in Figure 2). The tendency of valenced words to elicit higher arousal is consistent with a view that arousal is a measure of how strongly the perceived valence of the stimulus engages motivational aversive and appetitive systems and what level of energy the organism must mobilise to respond to the environmental need associated with the stimulus (Bradley, Codispoti, Cuthbert, & Lang, 2001; Duffy, 1951; Higgins, 2006; Lang, Bradley, & Cuthbert, 1990, among others). Furthermore, the boomerang shape of the functional relationship that we observe in our data is a hallmark of subjective judgements of affect and is robustly found in self-reports or aggregated ratings of, for instance, recent and distant emotional memories, current emotional states, emotionally laden and balanced pictures and, finally, words (Bradley & Lang, 1999; Kuppens et al., 2013).

Pitted against prior reports, our data patterns suggest that the word-stock of the English language provides more unique tools (i.e., word types) for the regions of affective space that are favoured in human subjective experience. Put differently, an increased communicative need to express certain experiences—the ones that link more extreme valence with higher arousal—appears to have prompted a more extensive creation and diversification of lexical items denoting those experiences. Remarkably, distributional patterns obtained from large corpora of English, which summarise the collective verbal behaviour of millions of speakers, dovetail perfectly with results of laboratory studies eliciting highly constrained responses (typically, ratings) to a hand-picked set of stimuli obtained from a much smaller number of participants.3 It is noteworthy that the convergence takes place even though our stimuli were selected without regard for emotionality and thus represent language in a naturalistic way, in stark contrast with prior studies offering either a carefully balanced representation of affective language or an overrepresentation of its extremes.

The availability of tools for expressing affect in the English language does not dictate how these tools are used: i.e., the distribution of word types over the affective space does not overlap with how often the types are selected for communication and how many tokens are associated with each word type. The preference in word tokens is for highly positive words (regardless of their level of arousal), as well as for extremely negative extremely arousing words, see Figure 3; the distribution is markedly different from the boomerang-like curve observed in word types, see Figure 2.

The body of findings presented earlier gives rise to a number of methodological and theoretical claims regarding the two novel distinctions that we made at the outset of the paper: valence vs. arousal and types vs. tokens. We discuss these in turn.

Valence vs. arousal

The methodological need for a joint consideration of the two key dimensions of affect (Boucher & Osgood, 1969) stems from the fact that one

---

3Kuppens et al. (2013) point out that the U- or a V-shaped relationship observed in ratings aggregated over hundreds or thousands individuals co-exists with very large individual variability in the relationship between arousal and valence. We are not able to test this between-levels difference as individual frequency distributions are not available to us.
positioning dimension plays a critical modulating role in the strength and direction of distributional biases shown by the other dimension. As Table 3 demonstrates, the advantage in the percent of positive words is attenuated and even significantly reversed as words become more arousing. A similar reversal of sign is observed in correlations of arousal with token frequency calculated per decile of valence. Thus, prior studies of the positivity bias essentially collapse a bi-dimensional distributional pattern onto a single dimension of valence. This is arguably more harmful for the boomerang-shaped functional curve observed in word types. When collapsed onto the valence dimension, many properties of the shape are not observable, including the presence of counter-directed slopes, the slopes’ relative magnitudes (reflecting potential asymmetry between the polarity of affect and its strength), or the inflexion point of the functional curve (as an index of true emotional neutrality; see Kuppens et al. 2013). A sole observable property of the boomerang shape is its very indirect characteristic, namely the number of words formed by a dissection of the boomerang at the midpoint of the valence scale, aggregated over levels of arousal. Similarly, focusing on positivity in the distribution of word tokens over the affective space misses a theoretically significant increase in token frequency associated with danger (negative arousing) words. To sum up, an accurate depiction of the spectrum of affect, as represented by lexical statistics of the English language, requires a bi-dimensional (or perhaps a multi-dimensional) perspective on both the words and the phenomena they denote.

Types vs. tokens

Separation of word types and tokens in our analyses as linguistically and psychologically independent indices of language use has received strong empirical support in the present data. Indeed, distributional patterns characterising the two types of linguistic units are highly dissimilar, from a boomerang shape in types to a concentration of tokens in the opposite bands and corners of the affective space. The type-token distinction also sheds light on, and enables, a revision of social factors proposed in the literature as causes of affective biases. These factors revolve around two statements about the emotional structure of the world and society: life contains more positive than negative events, concepts and objects (Augustine et al., 2011; Gable et al., 2000; Rozin et al., 2010) and humans consciously prefer to talk about the bright side of life to please the interlocutor and maintain more positive social interactions (Augustine et al., 2011). The present data considerably qualify these claims of a positive outlook and pro-social benevolence in communication.

As we argue earlier in this article, the word-stock of English is organised around subjective experience of affect. This implies that the bi-dimensional bias towards extremely valenced and arousing words and neutral calm words is not necessarily due to the prevalence of dangerous, exhilarating or even mundane phenomena in daily life. It is there because humans tend to preferentially assign these affective values to the spectrum of phenomena they encounter in their life: the bias is in the structure of subjective perception rather than in the structure of the world.

Furthermore, the tendency of English speakers to preferentially draw on all and any positive words and dangerous words in their communication from the word-stock which is not even especially diversified in all these regions of affect is symptomatic. We consider this bias in light of the long-standing proposal (Boucher & Osgood, 1969) that communicative behaviour is pro-social or benevolent in nature, where pro-sociality is broadly defined as a conscious choice of behaviour aiming at benefiting the recipient of the message (an individual interlocutor or a group). Importantly, however, pro-sociality is often equated with a tendency to preferentially look at the bright side of life (cf. Augustine et al., 2011; Boucher & Osgood, 1969). The observed data patterns—and especially a spike in token frequency for low-valence high-arousal words—corroborate the notion of pro-sociality, but not of the bias towards all matters positive. It stands to reason that language users benefit from communicating intensely about sources of danger, and
not only about sources of pleasure. Our data suggest then that conscious communicative behaviour, gauged by the choice of words and topics, is organised in a way that benefits recipients of written and spoken messages by providing a broader coverage of both phenomena that have a potential of a reward and that of a threat.

One explanation for the token-based bias towards these phenomena may come from the appetitive motivation to seek rewards and the fearful motivation to avoid threats (Bradley, 2000; Bradley & Lang, 2000). As argued in Lang, Bradley, and Cuthbert (1990), positive stimuli associated with usefulness for survival (including indices of sustenance, nurturance, and caregiving) or negative-arousing stimuli associated with danger have a privileged status in regulating hormonal control, engaging attention, and shaping cognitive processing (cf. Bradley, Codispoti, Cuthbert, & Lang, 2001; Wurm, 2007). Öhman and Mineka’s (2001) studies of fear in humans and primates further show that the threat-detection system is physiologically grounded, automatically activated and relatively impervious to cognitive control. We argue then that the statistical patterns observed in language use, i.e., word-token bias, faithfully demarcate the most salient regions of the affective space, the ones that show the most immediate and critical impact of emotion on physiological and cognitive processes. We add that the observed patterns do not allow for distinguishing between approach towards dangerous objects (with the purpose of attacking them) or avoidance of dangerous objects (with a purpose of escaping them) as a preferred behavioural strategy.

CONCLUSIONS

To summarise, a substantially large collection of emotional ratings has enabled us to identify and confirm distributional biases towards the usage of positive/negative and calm/arousing words in the English language. Statistical regularities of language use mirror both the emotional structure of the world and society and even more so the subjective emotional structure of a human being, with his or her primary motivations and behaviour in cognitive tasks influenced by affect. We argue for two factors that shape the structure of affect in society—as revealed via language. One is a prosocial benevolent communication strategy: we talk more about phenomena that can benefit our interlocutors by contributing to their survival and well-being, i.e., highly pleasurable and highly dangerous things. Another is prevalence and a broader diversification of lexical items expressing affective states that are most common in human subjective experience, with more extreme valence associated with higher arousal. In this sense, cumulative linguistic behaviour of vast collectives of language users replicates subtle experiential preferences observed in small groups of individuals. Finally, we argue that both factors are rooted in the fundamental motivational systems underlying emotional responses. As such, an accurate characterisation of affective biases provides a fuller understanding of how language, society, emotion and thought interrelate.

REFERENCES


APPENDIX

REANALYSIS OF KLOUMANN ET AL.’S (2012) DATA

Our observation of consistent correlations between word token frequency and word positivity across genres and language varieties (USA vs. UK), runs counter to a recent study by Kloumann et al. (2012) in which valence was found to only correlate very weakly with frequency in a few genres (Twitter and music lyrics) and not at all in others (Google Books and the New York Times). They concluded that valence and frequency were independent, a finding at odds not only with our results but also the reports of previously published work (see Introduction).

A close reading of Kloumann et al. (2012) methods reveals a few potential areas for concern. First, they did not employ any inclusion criteria with regards to the character strings they included in their sample, beyond the fact that the strings were among the top 5000 from each of their corpora. This led to an inclusion of multiple spelling variants (bdy, b-day and birthday), words with special characters (#music, #tcot), foreign words not borrowed into English (cf. Dutch hij “he” and zijn “to be”), alphanumeric strings (a3 and #p2) and others. The meaningfulness of happiness ratings for items such as these may be limited as a reflection of the emotional representation of the item in a typical speaker of English. (For comparison, all words in the stimulus list of Warriner et al. were identified as known by at least 70% of participants in another mega-study, Kuperman et al., 2012). Another consequence of Kloumann et al.’s decision to consider all character strings occurring in corpora “as is” is that word forms of the same lemma (e.g., walk, walks, walked, and walking as word forms of lemma walk, or table and tables as word forms of lemma table) are represented as emotionally discernible, independent language events. While such representation is logically possible, it is bound to represent a psychologically unlikely situation in which word forms of a lemma are each grounded in their own emotional experience and associated with values of positivity or arousal that are independent of those in other word forms related to the same lemma. We also note that the inclusion of word forms, instead of lemmas, in a word list, leads to inflation in the number of word types, and a lower token count, relative to lemma, for each specific word form: both implications of treating word forms independently are predicted to affect the type- and token-bias estimates.

We compared our results with Kloumann et al.’s in two ways. First we calculated Spearman correlations between our emotional ratings and ranked frequency for the overlapping words in each of their four corpora. Their correlations and ours are both reported in Table 4. They found the strongest relationship between emotion and frequency rank with their Twitter corpus, a relationship which is nearly equivalent to what we found in the overlapping words. However, they found only very weak relationships in the remaining three corpora where we found correlations 2 to 16 times stronger than theirs. We also divided each set of overlapping words into 10 deciles of frequency and plotted the distribution of valence in each (see Figure 4). In the plots, the proportion of words falling above the midpoint of the scale increases as frequency decile increases. Inset are plots showing how this change does not occur when Kloumann et al.’s full word list is used for each corpus along with their average happiness ratings and frequency ranks.
Table 4. Comparison of data from Kloumann et al. (2012) and Warriner et al. (2013)

<table>
<thead>
<tr>
<th></th>
<th>Twitter</th>
<th>Google Books</th>
<th>New York Times</th>
<th>Music Lyrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kloumann et al.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Words</td>
<td>5000</td>
<td>5000</td>
<td>5000</td>
<td>5000</td>
</tr>
<tr>
<td>% pos</td>
<td>72.0</td>
<td>78.8</td>
<td>78.4</td>
<td>64.1</td>
</tr>
<tr>
<td>Median V</td>
<td>5.54</td>
<td>5.64</td>
<td>5.56</td>
<td>5.34</td>
</tr>
<tr>
<td>Skewness V</td>
<td>-0.64</td>
<td>-0.89</td>
<td>-0.83</td>
<td>-0.44</td>
</tr>
<tr>
<td>ρ V</td>
<td>-0.103</td>
<td>-0.013</td>
<td>-0.044</td>
<td>-0.081</td>
</tr>
<tr>
<td>Warriner et al.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Words Overlap</td>
<td>2443</td>
<td>2704</td>
<td>2354</td>
<td>2458</td>
</tr>
<tr>
<td>% pos</td>
<td>73.3</td>
<td>72.9</td>
<td>74.3</td>
<td>66.7</td>
</tr>
<tr>
<td>Median V</td>
<td>5.73</td>
<td>5.59</td>
<td>5.68</td>
<td>5.57</td>
</tr>
<tr>
<td>Skewness V</td>
<td>-0.71</td>
<td>-0.66</td>
<td>-0.76</td>
<td>-0.53</td>
</tr>
<tr>
<td>ρ V</td>
<td>-0.101</td>
<td>-0.219</td>
<td>-0.187</td>
<td>-0.218</td>
</tr>
<tr>
<td>Median A</td>
<td>4.14</td>
<td>3.95</td>
<td>4.04</td>
<td>4.16</td>
</tr>
<tr>
<td>Skewness A</td>
<td>0.51</td>
<td>0.63</td>
<td>0.64</td>
<td>0.47</td>
</tr>
<tr>
<td>ρ A</td>
<td>0.034</td>
<td>-0.054</td>
<td>-0.020</td>
<td>0.044</td>
</tr>
</tbody>
</table>

Note: Note that in Kloumann et al. (2012), frequency estimates were provided as a ranking with lower numbers representing the highest frequency. As such, the correlations they report are opposite in sign to the correlations reported in other parts of this paper where frequency is reported as a log-transformed frequency count. For comparability, in the Warriner et al. portion of the table, we report correlations between Warriner et al.’s ratings and a similarly frequency ranking based on the frequencies provided in SUBTLEX.

Figure 4. Density plots for each of the corpora in Kloumann et al. (2012). In the main area, the overlapping words between Kloumann et al. and Warriner et al. (2013) are plotted based on ratings from Warriner et al. These words were divided into quartiles based on log Frequency, each quartile then being plotted separately—the solid line represents the lowest quartile and the dashed line the highest quartile. The inset plots use the full data-set from Kloumann et al. along with their frequency ranks and happiness ratings.