Language-based Valence and Arousal Expressions between the United States and China: a Cross-Cultural Examination

Young-Min Cho[1] **Dandan Pang**[2] **Stuti Thapa**[3] **Garrick Sherman**[1]
Lyle Ungar[1] **Louis Tay**[4] **Sharath Chandra Guntuku**[1]
[1]University of Pennsylvania [2]Bern University of Applied Sciences
[3]University of Tulsa [4]Purdue University
{jch0,ungar,sharathg}@seas.upenn.edu, dandan.pang@bfh.ch, stt0912@utulsa.edu
garricktsherman@gmail.com, stay@purdue.edu

Abstract

While affective expressions on social media have been extensively studied, most research has focused on the Western context. This paper explores cultural differences in affective expressions by comparing valence and arousal on Twitter/X (geolocated to the US) and Sina Weibo (in Mainland China). Using the NRC-VAD lexicon to measure valence and arousal, we identify distinct patterns of emotional expression across both platforms. Our analysis reveals a functional representation between valence and arousal, showing a negative offset in contrast to traditional lab-based findings which suggest a positive offset. Furthermore, we uncover significant cross-cultural differences in arousal, with US users displaying higher emotional intensity than Chinese users, regardless of the valence of the content. Finally, we conduct a comprehensive language analysis correlating n-grams and LDA topics with affective dimensions to deepen our understanding of how language and culture shape emotional expression. These findings contribute to a more nuanced understanding of affective communication across cultural and linguistic contexts on social media.[1]

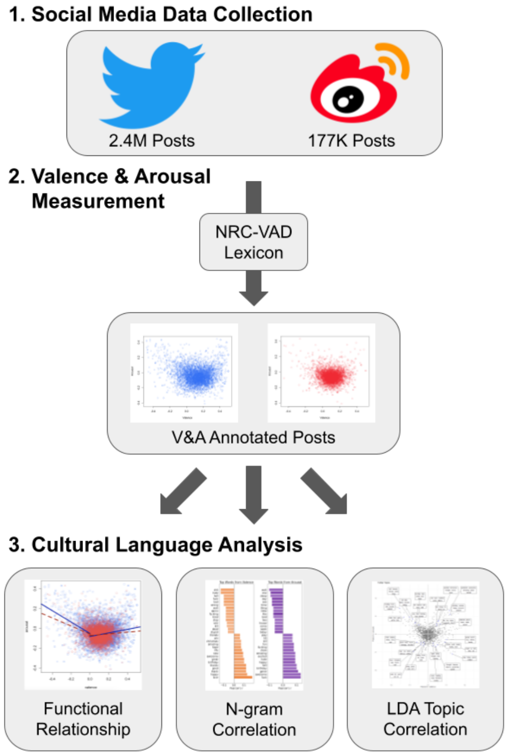


Figure 1: The analysis pipeline of this paper compares cultural differences in affective expressions using large-scale social media data. We examine the functional relationship between valence and arousal and explore the differences through language analysis methods.

1 Introduction

Subjective expressions of affect (how we feel) play a crucial role in understanding learning outcomes in individuals (Hourihan et al., 2017), their perceptions (Gorn et al., 2001), well-being (Xu et al., 2015), and mental and physical health (Cohen and Pressman, 2006). Multiple theoretical and empirical works have, therefore, examined the underlying dimensions of affect and their relationships. While there are several models of affective structure, Russell's two-dimensional circumplex model is the most widely recognized, where orthogonal valence (pleasant to unpleasant) and arousal (high to low activation) are represented as the horizontal and vertical axes[2] (Russell, 1980; Yik et al., 1999).

Understanding the functional relationship between valence and arousal in this two-dimensional space is of empirical, psychometric, and theoretical interest. Among various models from previous studies (Ortony et al., 1990; Lang, 1994), a "V-shaped" relationship, where arousal is a function of valence, is one of the most widely tested and accepted (Kuppens et al., 2013; Cacioppo and Gardner, 1999). Arousal is shown to be directly re-

[1]For future research, we release our dataset at https://github.com/JeffreyCh0/X_Weibo_Valence_Arousal

[2]See Russell's two-dimensional circumplex model in Figure 6.

Findings of the Association for Computational Linguistics: NAACL 2025, pages 5601–5615

lated to the intensity of positive or negative valence with a positivity offset and a negativity bias, with varying levels of cross-cultural support (Kuppens et al., 2017a).

While the affective structure and the valence-arousal relationship are often considered universal, most previous studies focus exclusively on Western samples, overlooking cross-cultural heterogeneity (Tsai et al., 2006a). Different cultures value emotions (ideal affect) uniquely and adhere to distinct standards for emotional expression (Matsumoto, 1990). For instance, Americans tend to associate enthusiasm (high arousal) with positive valence, while Asians often prefer quietness (low arousal) (Tsai et al., 2006a). Although some studies have highlighted affective differences between Western and Eastern cultures, these are typically based on small, lab-based samples (Kuppens et al., 2017b), which may be biased by self-reporting and recall issues (Tarrant et al., 1993; Winograd and Neisser, 2006).

To address these limitations, researchers advocate for studies that go beyond self-reports and focus on behaviors (Baumeister et al., 2007), especially on social media. Social media data offer a naturalistic and ecological setting to capture individuals' emotions and, despite potential social desirability bias, have been shown to reliably estimate well-being (Jaidka et al., 2020; Liou et al., 2023), sentiment (Preoţiuc-Pietro et al., 2016), and personality (Schwartz et al., 2013a; Havaldar et al., 2024).

This paper examines the cross-cultural difference on the affective expressions with functional relationship between valence and arousal by analyzing natural language expressions from Twitter/X (geolocated to the US) and Sina Weibo (in Mainland China) posts. The pipeline of our work is shown in Figure 1. The study has three key contributions:

- We evaluate functional representations between valence and arousal on large-scale social media data, identifying a negative offset in contrast to previous lab-based studies.
- We demonstrate cross-cultural differences in valence and arousal, showing that US users exhibit stronger emotional intensity (higher arousal) than Chinese users across both positive and negative valence.
- We employ comprehensive language analysis, correlating n-grams and LDA topics with valence and arousal, providing insights into the functional relationship and cultural divergence in affective expression.

2 Methods

2.1 Data Preparation

Our data consist of public messages posted on Weibo and Twitter. The content and behavior variations on Weibo and Twitter have been studied in different contexts (Ma, 2013; Lin et al., 2016). While working with non-random, non-representative samples poses challenges, social media posts can still reveal psychological traits, demographics (Sap et al., 2014; Zhang et al., 2016), location (Salehi et al., 2017; Zhong et al., 2015), and mental health (Guntuku et al., 2019c; Tian et al., 2018).

To collect Twitter data, we used the survey platform Qualtrics, which included demographic questions such as gender and age. Participants from the US shared their Twitter handles after completing the survey. Users were compensated for their time, and consented to share their Twitter posts. There were 3,113 Twitter users, with around 3.6 million posts until 2016.

Weibo, unlike Twitter, does not offer an API tool for obtaining random samples over time. So, starting with a random set of individuals from a public dataset (Guntuku et al., 2019a), Weibo posts were gathered using a breadth-first search method on users[3]. We obtained over 29 million posts from 2014 from 859,054 people on Weibo. Gender and age were collected from self-reported demographic information on their Weibo profile. Subsetting to users posted more than 500 words and with a reasonable self-reported age (<100 years) and gender, the dataset consisted of 668,257 Weibo posts from 8,731 users. 500 words were found to be the minimum threshold to obtain reliable psychological estimates from individuals' language (Eichstaedt et al., 2021; Jaidka et al., 2018).

Based on the gender and age distribution of Weibo and Twitter users, we built propensity-score-based matched samples, resulting in 2,191 users each on both platforms with at least 500 words[4] (Rosenbaum and Rubin, 1983). These matched

[3]We started with a random set of users, and expanded to all their friends (bidirectional, similar to followers + following), and we repeated the process.

[4]We use propensity score matching for its nuanced handling of continuous variables and its allowance for quantitative assessment of covariate balance between matched groups.

users had 2.4 million posts on Twitter and 177,042 posts on Weibo. In our matched dataset 67.1% self-reported as being female and 32.9% as male, and the mean age was 26.9 (s.d. 8.8). On Twitter, there were on average 15.6 (s.d. 2.8) words per user and Weibo had 57.3 (s.d. 15.4) words per user. The differences in word counts are driven by the lack of character limits to posts on Weibo.

To eliminate the confounds of bilingualism (Fishman, 1980), we retain only English posts on Twitter and Mandarin posts on Weibo by using langid (Lui and Baldwin, 2012). Re-tweets are also removed from both datasets ('RT @USERNAME:' on Twitter and '@USERNAME//' on Weibo). Weibo posts were split into tokens using THULAC (Li and Sun, 2009) while Twitter posts were segmented using happierfuntokenizing (DLATK/happierfuntokenizing, 2017) due to their ability to handle emoticons and other social media slang. To eliminate uncommonly used words (outliers), we filtered words with different frequency thresholds for each platform. Words used by fewer than 0.1% of the total posts on Twitter and 0.5% on Weibo were removed from the analysis. Most words are seldom used in language, as they follow a Zipfian distribution. By removing these words, we ensure that the language insights from our research can be generalized to out-of-sample cases.

2.2 Valence & Arousal Measurement

The circumplex and vector models of emotion have been broadly used for representing affective states[5] (Russell, 1980; Bradley et al., 1992). In these two-dimensional models, valence is the x-axis, expressing pleasantness and unpleasantness, attractiveness and aversiveness, joy, and sorrow (Frijda, 1986). Arousal is the y-axis, describing the degree of wakefulness, boredom, excitement, and calm. These models allow any affective state, emotion, word, or expression to be represented as a point in the space, regardless of the difference in language, country, or culture.

We measure valence and arousal using a validated data-driven lexicon generated based on the circumplex model in both English and Mandarin. We used NRC Valence, Arousal, and Dominance (NRC-VAD) Lexicon (Mohammad, 2018a) for Twitter data and its translated version for Weibo data. NRC-VAD consists of valence and arousal weights for more than 20,000 words in English and shows a "V-shaped" relationship between two dimensions: extremely positive or negative valence is usually paired with high arousal, while calmness matches low arousal. We subtract 0.5 from all scores to make them zero-centered.

Multilinguality is another reason to choose NRC-VAD as our valence-arousal measurement lexicon. There are over 100 languages available for NRC-VAD (August 2022), and the authors claim that most affective norms are stable across languages. Since an original-translated term pair has the same scores, this lexicon avoids the annotator agreement and scale-matching issue, which are common problems using two different lexica over two languages.

We calculate the valence and arousal scores for each post on Twitter and Weibo using NRC-VAD lexica. For each post, we sum the result of item-wise multiplication of relative word frequency within the post, and the corresponding valence or arousal score for the word. In detail, we follow the formula:

$$valence_m = \sum_{w \in W_m} valence_w \cdot \frac{freq_m(w)}{W_m} \quad (1)$$

Where m represents a post, w is a word, $freq_m(w)$ is the frequency of word w in post m, $valence_m$ and $valence_w$ are the valence scores for post m (for annotation) and word w (from the NRC-VAD lexicon), respectively, and W_m is the total number of words in post m. Arousal is calculated in a similar manner.

2.3 Evaluation of Functional Relationship

Kuppens et al., 2013 showed six possible functional relationships between valence and arousal. These models are independence (Model 1), Linear Relation (Model 2), Symmetric V-Shaped Relation (Model 3), and Asymmetric V-Shaped Relations, including asymmetric interception (Model 4), asymmetric slope (Model 5), and asymmetric interception and slope (Model 6). The models' functional representations are shown below[6]:

$$A_m = \begin{cases} \beta_0 + \epsilon_m & \text{(Model 1)} \\ \beta_0 + \beta_1 V_m + \epsilon_m & \text{(Model 2)} \\ \beta_0 + \beta_1 |V_m| + \epsilon_m & \text{(Model 3)} \\ \beta_0 + \beta_1 |V_m| + \beta_2 I_m + \epsilon_m & \text{(Model 4)} \\ \beta_0 + \beta_1 |V_m| + \beta_3 I_m |V_m| + \epsilon_m & \text{(Model 5)} \\ \beta_0 + \beta_1 |V_m| + \beta_2 I_m + \beta_3 I_m |V_m| + \epsilon_m & \text{(Model 6)} \end{cases}$$

[5] See Russell's two-dimensional circumplex model in Figure 6.

[6] See full examples on our dataset on Figure 10.

Where A_m and V_m are short for $Arousal_m$ and $Valence_m$, arousal and valence scores for the post m, I_m denotes a dummy variable that indicates whether $Valence_m$ is positive($I_m = 1$) or negative($I_m = 0$). Each model is tested with a within-person intercept and slope.

We use Akaike Information Criterion (AIC; Bozdogan, 1987), Bayesian Information Criterion (BIC; Schwarz, 1978), posterior probability, and Conditional R^2 for model selection. AIC is defined as $AIC = 2 \cdot k - 2 \cdot \ln(\hat{L})$, and BIC has the following format: $BIC = -2 \cdot \ln(\hat{L}) + k \cdot \ln(N)$, where $\hat{L}$ is the maximized value of the likelihood function of the model, k is the number of parameters, and N is the number of observations. One advantage of using BIC is that it can be used to approximate posterior probability for each model:

$$P(model_i|data) = \frac{exp(-0.5BIC_i)}{\sum exp(-0.5BIC_i)} \quad (2)$$

While applying the six models, we use mixed effects models to fit the datasets. We assume there is a fixed relationship between valence and arousal across all posts, while the average level of arousal may vary from user to user. The regression models can correctly represent the relationship between the two variables by setting within-person differences as the random effect. In the model comparison, to cover both fixed and random effects, we use conditional R^2 to represent the proportion of variance explained by the entire model.

2.4 Social Media Language Analysis

Feature Extraction We extract two open-vocabulary features from Twitter and Weibo: n-grams and topics. N-grams help capture common word patterns and phrase structures that reflect how emotions are expressed in everyday language, allowing us to identify culturally specific linguistic cues tied to affective dimensions. Meanwhile, Latent Dirichlet Allocation (LDA, Blei et al., 2003) uncovers underlying topics within the text, revealing thematic contexts that influence emotional expression. We choose LDA for topic modeling because it has better explainability and computational efficiency than other modern models like Top2Vec and BERTopic (Angelov, 2020; Grootendorst, 2022). By analyzing both n-grams and topics, we can better understand the interplay between language, culture, and emotion, providing a richer, data-driven perspective on cross-cultural affective communication.

We collect contiguous sequences of one or two words (1-2 grams, Kern et al., 2014; Andrew Schwartz et al., 2013) with pointwise mutual information (PMI = 3; Church and Hanks, 1990). This resulted in unique unigrams and bigrams set of 10,477 for Weibo and 12,798 for Twitter. We extracted the normalized distribution of the n-grams for each post in the Weibo and Twitter datasets. We then used 2,000 topics generated using LDA as the second feature to represent users' language in our Twitter and Weibo datasets (Schwartz et al., 2013b). We utilized topics generated on much larger datasets to favor high diversity and coverage. 2,000 English topics generated a corpus of approximately 18 million Facebook updates with alpha set to 0.30 to favor fewer topics per document. These have been shown to perform well across multiple platforms (Eichstaedt et al., 2015). 2,000 Mandarin topics were generated on 29 million Weibo posts with similar parameters set in Mallet (McCallum, 2002). Inherently, each topic is realized as a set of words with probabilities. Every post is thus scored in terms of its probability of containing each of the 2,000 topics, $p(topic, post)$, which is derived from their probability of containing a word, $p(word|post)$, and the probability of the words being in given topics, $p(topic|word)$.

Differential Language Analysis To understand the functional relationship and cultural difference in valence and arousal, we utilized ordinary least squares (OLS) regression to model valence and arousal on post level, controlling for gender and age by matching these variables between the Twitter and Weibo samples. The inputs for the regression were different language features independently extracted from social media posts - n-grams and topics; and outputs were valence and arousal, each of which was constructed using a separate OLS model. From OLS regression, we extracted coefficients to represent Pearson correlation coefficients for each feature dimension[7]. To correct for multiple comparisons and control the false discovery rate in multiple hypothesis testing, we applied the Benjamini-Hochberg p-correction (Benjamini and Hochberg, 1995). We considered correlations meaningful if they met the threshold of $p < .05$.

[7] For details of how Pearson r is calculated, see Appendix A

$$Arousal_m = \beta_0 + \beta_1 |Valence_m| + \beta_2 I_m + \beta_3 I_m |Valence_m| + \varepsilon_m$$

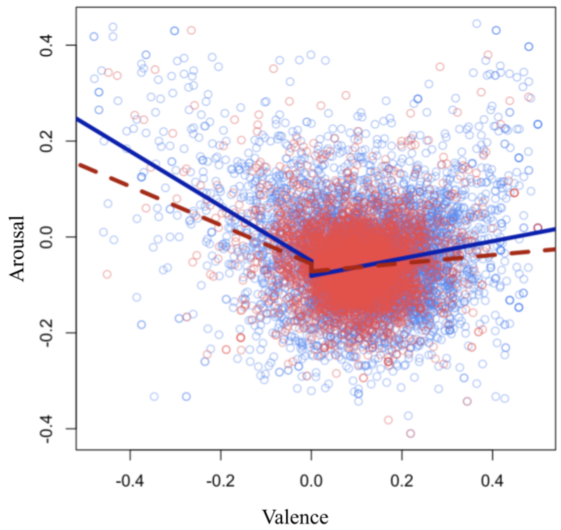


Figure 2: Scatter plots of valence (x-axis) and arousal (y-axis) of Twitter (blue) and Weibo (red) posts. The lines of best fit for Model 6's function are appended to the plot(Twitter: solid line, Weibo: dashed line). The model is tested with within-person intercept and slope.

The formula for n-gram models are:

$$valence_m \sim \sum_{n \in 1-2gram} a_n \cdot \frac{freq_m(n)}{N_m} + \varepsilon \quad (3)$$

where m is a post, n is an n-gram, $freq_m(n)$ is the frequency of n in m, a_n is coefficients and N_m is a total number of n-grams in the post.

The formula for LDA topic models are:

$$valence_m \sim \sum_{t \in Topics} a_t \cdot P(t|m) + \varepsilon \quad (4)$$

where $P(t|m)$ is the probability of m belonging to topic t. Models for arousal can be expressed in the same fashion.

3 Results

Warning: The following section contains swear words.

3.1 Valence-Arousal Functional Relationship

The full comparison between 6 models are shown in Figure 10. Among the different models we tested across Twitter and Weibo data, Model 6 with within-person intercept and slope best fit with the lowest AIC and BIC, and highest Conditional R^2. Within-person models were also significantly different from the models without within-person effects.

As shown in Figure 2, the presence of an asymmetric V-shape in the data, including a negativity bias and negativity offset, was confirmed in the models on both Twitter and Weibo data. Compared with Weibo, Twitter shows a larger intercept gap (Twitter: $\beta_2 = -0.031$; Weibo: $\beta_2 = -0.016$). The intensity of emotion gets significantly stronger with higher positivity/negativity. This conclusion is consistent with both Twitter and Weibo, with the smallest BIC values in Model 6, characterized by a V shape (Twitter: $\beta_1 = 0.573$, Weibo: $\beta_1 = 0.404$) and negativity bias (Twitter: $\beta_3 = -0.392$, Weibo: $\beta_3 = -0.318$). The Twitter model has a steeper slope on both positive and negative valence compared to Weibo (Twitter: $\beta_1 = 0.573$, $\beta_3 = -0.392$; Weibo: $\beta_1 = 0.404$, $\beta_3 = -0.318$).

3.2 Social Media Language Analysis

To uncover the content differences in emotional expression across cultures, we utilized differential language analysis to obtain the most correlated n-grams and topics in each platform. Figure 3 shows the top significantly correlated words and phrases with valence and arousal in both platforms. On the dimension of valence, Twitter users tended to use words conveying superlatives ('great', 'awesome', 'amazing') and festive celebrations ('birthday', 'Christmas', 'new', 'win') in expressing positive valence, while profanity ('shit', 'fuck'), negation ('hate', 'bad', 'wrong') and discomfort ('wait', 'tired', 'stop') were indicative of negative valence. Conversely, Weibo users commonly employed terms related to personal affect ('like', 'love', 'happiness') and emojis ('oh', 'heart') when expressing positive valence, whereas words indicative of negation ('no') and sorrow ('sad', 'cry') are prevalent

Dataset	Model	AIC	BIC	PostP	R^2
Twitter	Model 1	-3.752×10^6	-3.752×10^6	0	0.015
	Model 2	-3.790×10^6	-3.790×10^6	0	0.030
	Model 3	-3.833×10^6	-3.833×10^6	0	0.050
	Model 4	-4.001×10^6	-4.001×10^6	0	0.113
	Model 5	-4.048×10^6	-4.048×10^6	0	0.135
	Model 6	$\mathbf{-4.060 \times 10^6}$	$\mathbf{-4.060 \times 10^6}$	**1**	**0.139**
Weibo	Model 1	-4.155×10^5	-4.155×10^5	0	0.021
	Model 2	-4.162×10^5	-4.161×10^5	0	0.025
	Model 3	-4.172×10^5	-4.171×10^5	0	0.030
	Model 4	-4.219×10^5	-4.219×10^5	0	0.057
	Model 5	-4.247×10^5	-4.246×10^5	0	0.082
	Model 6	$\mathbf{-4.251 \times 10^5}$	$\mathbf{-4.250 \times 10^5}$	**1**	**0.084**

Table 1: Results of fitting 6 different models on Twitter and Weibo dataset. AIC is Akaike Information Criterion, BIC is Bayesian Information Criterion (the lower the better fit), PostP is posterior probability, R^2 is Conditional R^2.

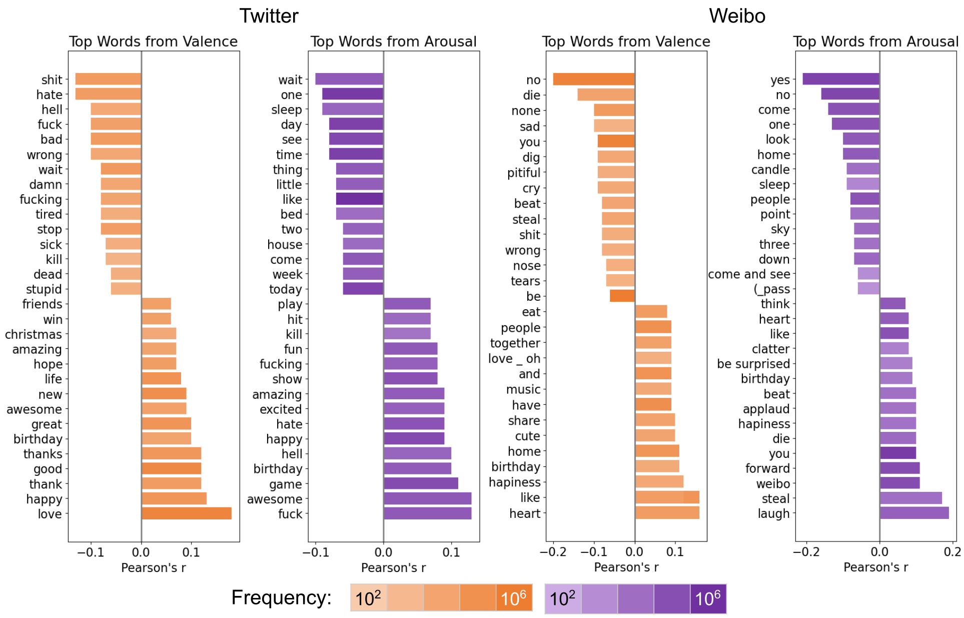


Figure 3: Words and phrases associated with valence and arousal on Twitter and Weibo (translated) from the top 15 phrases for effect strength (Pearson r), colored by frequency. Statistically significant ($p < .05$, two-tailed t-test, Benjamini-Hochberg corrected).

in expressing negative valence. On the dimension of arousal, Twitter users expressed profanity ('shit', 'fuck') and interpersonal expressions ('awesome', 'amazing') for high arousal while using terms indicating low activities ('sleep', 'bed') and time-oriented description ('today', 'week', 'day', 'time') for low arousal. In contrast, Weibo users predominantly utilized positive emojis('steal-laugh', 'applaud') to convey high arousal, while employing affirmation ('yes'), negation ('no'), and sharing aspects of daily life ('home', 'sleep') to express low arousal. The version of Figure 3 without using words in NRC-VAD is shown in the Appendix.

We further compare Twitter and Weibo's LDA results for topics in Figure 4 and 5. Twitter users had relaxing weekend ('weekend', 'awesome', 'amazing', 'great', 'retreat'), celebration of events ('birthday', 'wishes', 'happy', 'present', 'wished'), luck and achievement ('win', 'won', 'contest', 'prize', 'lottery') for positive valence high arousal. Conversely, Weibo users discussed affectionate bonding ('love', 'hopeless', 'willing', 'protective', 'friendly', where hopeless means love in deep) to express their feelings, particularly in the context of festivals and celebrations ('new year', 'red envelope') and interests in celebrities and TV shows ('celebrities','singer'). For positive valence low arousal, Twitter users usually talked about relaxing routines ('day', 'today', 'good', 'chilled') and sleep ('night', 'sleep', 'tonight', 'rest', 'hoping'). Besides, Weibo users shared family reunion ('home', 'return', 'mother', 'family', 'new year') and savory cuisines ('dish', 'meat', 'delicious', 'soup', 'dish'). When expressing strong negative feelings, Twitter users mainly used profanities ('fucking', 'fuck', 'shit', 'pissed', 'bullshit') to convey intense emotions, while Weibo users discussed law enforcement and criminal investigation ('police', 'crime', 'suspect', 'caught', 'case'). Additionally, Weibo discussions on negative high arousal included the use of emojis ('sweat') and negative emotions ('shocking', 'hurt', 'give up'). Concerning negative valence low arousal, Twitter users usually showed personal negative feelings like tiredness ('tired', 'sleepy', 'sleep', 'sooo', 'ugh'), engaged in discussions about daily activities ('hair', 'cut', 'short', 'haircut', 'cutting') and mentioned words related to time ('hour', 'minute'). Similarly, Weibo users also mentioned sleep ('sleep', 'awake', 'bed').

4 Discussion

This paper examined the functional relationship between valence and arousal based on large-scale social media texts across the United States and China. Our findings suggest that public affective expressions replicate the asymmetrical affective V-shaped relationship but with a negativity bias (negative feelings increase more strongly than positive feelings with increasing arousal) and negativity offset (feelings of arousal are higher at low negative va-

Figure 4: Topics associated with valence and arousal on Twitter, sorted by effect size (Pearson r). Each point is a topic, and statistically significant topics ($p < .05$, two-tailed t-test, Benjamini-Hochberg corrected) are shown in dark gray. The X-axis is the Pearson r with valence and the Y-axis with arousal. The top 5 words in each topic are shown.

lence levels than positive valence). In addition, the arousal and valence slope was steeper for Twitter users than for Weibo users.

One of the major findings in our study is that the American participants had stronger negativity bias and overall had higher arousal with higher positive and negative valence compared to Chinese participants. This is consistent with past findings on West-East distinction in emotional arousal and aligns with Hofstede's Individualism vs. Collectivism dimension: in Western or individualist culture, high-arousal emotions are valued and promoted more than low-arousal emotions, while in Eastern or collectivist culture, low-arousal emotions are valued more than high-arousal emotions (Lim, 2016; Hofstede, 2011). This can be attributed to the fact that individualistic culture encourages expressive independence and the externalization of personal emotions while collectivist culture values social harmony and group cohesion. Even in traditional Asian medicine, there is an assumption that excessive emotional expression can be harmful and cause diseases, whether it is positive or negative emotions (Lim et al., 2008). Our findings confirmed that Chinese users of Weibo express lower arousal levels for both negative and positive emotions.

Content analyses of the findings suggested that Chinese participants displayed less high arousal positive affect emotional behavior than their American counterparts. This is consistent with past findings that there seems to be a general preference in the West for high-arousal positive states like excitement or enthusiasm (Sommers, 1984). At the same time, people in the East generally prefer low-arousal positive affective states like calm or peacefulness (Tsai, 2007). Moreover, we saw Twitter users using more explicit excitement-focused terms such as awesomeness, while Weibo users tended to express positive emotions more implicitly, e.g., emojis. This is consistent with findings that the communication style of East Asian language

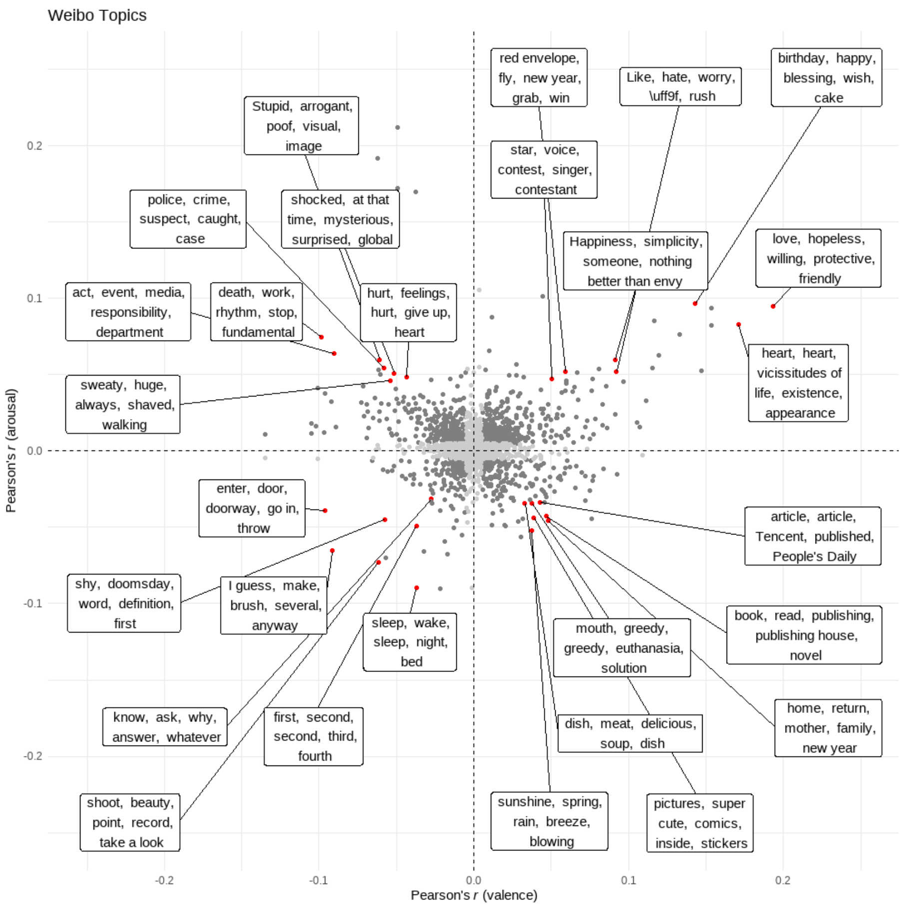


Figure 5: Topics associated with valence and arousal on Weibo, sorted by effect size (Pearson r). Each point is a topic and statistically significant topics ($p < .05$, two-tailed t-test, Benjamini-Hochberg corrected) are shown in dark gray. The X-axis is the Pearson r with valence and the Y-axis with arousal. English translations of the top 5 words in each topic are shown.

communities tends to be more indirect than that of their Western counterparts (Fong, 1998; Gudykunst et al., 1988; Neuliep, 2012).

Similarly, past literature suggests that high-arousal emotions serve as an effective means of influencing others in the West (Tsai, 2007), while low-arousal emotions serve as an effective means of adjusting and conforming to others in the East (Markus and Kitayama, 1991). We found that low arousal emotions in Weibo were used to create a sense of comfort and connection through themes related to nature, family, daily life activities, and light-hearted entertainment. On the high arousal-high positive affect sphere, Twitter users celebrated more personal events, while Weibo users talked more excitedly about celebrities and current events. Therefore, it is likely that while the ideal affect preference translates into affective expressions about personal experiences in the East, discussion of media culture is exempt from such norms: for instance, while it may be frowned upon to act too excited about personal events, the same restrictions are not in place when expressing excitement about celebrities and cultural events. As such, our findings provide a novel insight into our understanding of norm differences in affective expression in East vs West.

Similarly, looking at the difference in negativity bias for Twitter and Weibo, while Twitter users use profanity primarily, Weibo users tend to use words with much lower intensity, confirming the assumption that Chinese users try to avoid expressing extreme emotions. Note that although Weibo has censorships, it does not include profanity filters.

One surprising finding in our study was that we did not find a positivity offset. We instead found a negativity offset for both American and Chinese participants. The theoretical explanation for the positivity offset (and negativity bias) comes from the Evaluative Space Model (ESM; Cacioppo et al., 1999; Norris et al., 2010), which proposed that positive and negative affect have different arousal functions and predicts greater positive than negative affect at low levels of affective input. The adaptive reason for the offset was hypothesized to encourage approaching novel stimuli in low-threat conditions. However, our finding suggests this may not translate to public affective behavior, particularly on social media. It suggests that people on both Twitter and Weibo are more likely to approach neutral stimuli in negative terms while simultaneously having stronger negative reactions to higher arousal events. Therefore, our studies elucidate how certain theories of affect may not explain affective behavior universally, partly because of the contexts not considered in said theories.

5 Conclusion

This study highlights the importance of studying public emotional behavior and how it is distinguished from self-reported findings. Our findings could confirm some theoretical assumptions in traditional self-report research by adding new empirical evidence when applied to public emotional behavior. Future research looking at individual self-reports and public behavior can help us understand what these differences can represent at the individual level.

6 Limitations

Platform Issue: Even though Twitter and Weibo are comparable in usage (Li et al., 2020; Guntuku et al., 2019b) and have not been shown to have significant differences in predicting individual states (Gao et al., 2012), data from other platforms such as WeChat and RenRen in China and Facebook in the US have not been included in this study due to access constraints. Emojis are a significant contributor to affective expressions (Li et al., 2019); however, we did not include them in this study due to differences in encodings while collecting the data making it infeasible for us to parse them accurately. Further, social media users are non-representative of the general population, and the participants in this study are non-random and convenient samples.

Fine-grained Emotion: Our focus in this paper was to compare the expressions of valence and arousal across two different cultures building upon rich cross-cultural psychological studying the difference in valence and arousal (Lim, 2016; Kuppens et al., 2017a). Although considering fine-grained emotions could make the analysis multi-dimensional, it will make the results less reliable. Moreover, each of the fine-grained emotions could be represented in the valence and arousal circumplex (Jefferies et al., 2008; Mohammad, 2018b).

Subcultural Variance: We acknowledge the existence of subcultural variances, such as those among various ethnicities, provinces, and counties in China and the US. For example, minority students, including Tibetan and Mongolian, tend to experience more negative emotions and are less inclined to adopt emotion regulation strategies compared to Han students (Lü and Wang, 2012). Within the United States, European Americans show a greater motivation to engage in hedonic emotion regulation than their Asian American counterparts (Miyamoto et al., 2014).

However, despite these nuances, the macro-cultural differences between East and West remain significant enough to warrant a comparative analysis (Lu et al., 2001; Tsai et al., 2006b). Focusing on broader cultural variation, our investigation emphasizes the pronounced disparities between the US and Chinese cultural contexts by using social media posts. These disparities are substantial and provide a robust framework for comparative analysis. This macro-level perspective is not to negate the relevance of subcultural variances but to highlight the overarching patterns that emerge when contrasting Eastern and Western cultures by using two large countries that have a large variation in Hofstede's cultural dimensions, for instance, as examples. By situating our work within this broader context, we aim to contribute to a more comprehensive understanding of how culture influences emotional expression. In the discussion, we will add the above language to acknowledge the complex variety of cultural diversity that exists within and across national borders.

Translation: Lexica need to be adapted to the cultures to measure psychological phenomenon accurately. We tried using Chinese valence-arousal words (CVAW, Lee et al., 2022). However, we did not proceed further as the methods of building NRC-VAD (for English) and CVAW (for Chinese) lexica were different and could cause misalignment. We wanted to control for such differences by choosing a lexicon that has sufficient coverage while also being used in multiple prior works across both languages (Wenjia et al., 2023; Mohammad, 2016; Li et al., 2019; Chen et al., 2019; Das and Dutta, 2021). Further, with over 20K entries, NRC-VAD is the largest manually created emotion lexicon that has translations across several languages.

Censorship: We acknowledge that censorship on Sina Weibo is a challenge. Despite this issue, Weibo has been successfully used across multiple studies to understand different psychological outcomes (e.g. affect, stress, depression; Pan et al., 2021; Tang et al., 2019; Li et al., 2014)

7 Ethics

The University of Pennsylvania's Institutional Review Board declared this project exempt (IRB protocol # 829811).

This study, focusing on the cultural differences in affective expressions between Twitter users in the United States and Sina Weibo users in China, raises several ethical considerations:

1. Data Privacy and Anonymity: The research analyzes social media posts from Twitter and Sina Weibo. It is important to ensure that individual users' privacy is respected. All data extracted from these platforms is anonymized by removing personally identifiable information.

2. Cultural Sensitivity and Bias: Given the cross-cultural nature of the study, it is critical to approach the analysis with cultural sensitivity. Researchers must be aware of and mitigate any biases arising from their cultural backgrounds or perspectives. This includes being mindful of how cultural contexts influence affective expressions and the interpretation thereof.

3. Representation and Generalization: Care should be taken to avoid over-generalizing the findings. The study's results are based on specific social media platforms and may not represent the broader United States and China populations.

References

H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin E P Seligman, and Lyle H Ungar. 2013. Personality, gender, and age in the language of social media: The Open-Vocabulary approach. *PLoS One*, 8(9):e73791.

Dimo Angelov. 2020. Top2vec: Distributed representations of topics. *arXiv preprint arXiv:2008.09470*.

Roy F Baumeister, Kathleen D Vohs, and David C Funder. 2007. Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on psychological science*, 2(4):396–403.

Yoav Benjamini and Yosef Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

Hamparsum Bozdogan. 1987. Model selection and akaike's information criterion (aic): The general theory and its analytical extensions. *Psychometrika*, 52(3):345–370.

M M Bradley, M K Greenwald, M C Petry, and P J Lang. 1992. Remembering pictures: pleasure and arousal in memory. *J. Exp. Psychol. Learn. Mem. Cogn.*, 18(2):379–390.

John T Cacioppo and Wendi L Gardner. 1999. EMOTION.

John T Cacioppo, Wendi L Gardner, and Gary G Berntson. 1999. The affect system has parallel and integrative processing components: Form follows function. *Journal of personality and Social Psychology*, 76(5):839.

Guanhang Chen, Lilin He, and Konstantinos Papangelis. 2019. Sentimental analysis of chinese new social media for stock market information. In *Proceedings of the 2019 the International Conference on Pattern Recognition and Artificial Intelligence*, pages 1–6.

Kenneth Church and Patrick Hanks. 1990. Word association norms, mutual information, and lexicography. *Comput. Linguist.*, 16(1):22–29.

Sheldon Cohen and Sarah D Pressman. 2006. Positive affect and health. *Curr. Dir. Psychol. Sci.*, 15(3):122–125.

Subasish Das and Anandi Dutta. 2021. Characterizing public emotions and sentiments in covid-19 environment: A case study of india. *Journal of Human Behavior in the Social Environment*, 31(1-4):154–167.

Johannes C Eichstaedt, Margaret L Kern, David B Yaden, H A Schwartz, Salvatore Giorgi, Gregory Park, Courtney A Hagan, Victoria A Tobolsky, Laura K Smith, Anneke Buffone, Jonathan Iwry, Martin E P Seligman, and Lyle H Ungar. 2021. Closed- and open-vocabulary approaches to text analysis: A review, quantitative comparison, and recommendations.

Johannes C Eichstaedt, Hansen Andrew Schwartz, Margaret L Kern, Gregory Park, Darwin R Labarthe, Raina M Merchant, Sneha Jha, Megha Agrawal, Lukasz A Dziurzynski, Maarten Sap, et al. 2015. Psychological language on twitter predicts county-level heart disease mortality. *Psychological science*, 26(2):159–169.

Joshua A Fishman. 1980. Bilingualism and biculturism as individual and as societal phenomena.

Mary Fong. 1998. Chinese immigrants' perceptions of semantic dimensions of direct/indirect communication in intercultural compliment interactions with north americans. *Howard journal of Communication*, 9(3):245–262.

Nico H Frijda. 1986. *The Emotions*. Maison des Sciences de l'Homme.

Qi Gao, Fabian Abel, Geert-Jan Houben, and Yong Yu. 2012. A comparative study of users' microblogging behavior on sina weibo and twitter. In *User Modeling, Adaptation, and Personalization: 20th International Conference, UMAP 2012, Montreal, Canada, July 16-20, 2012. Proceedings 20*, pages 88–101. Springer.

Gerald Gorn, Michel Tuan Pham, and Leo Yatming Sin. 2001. When arousal influences ad evaluation and valence does not (and vice versa). *J. Consum. Psychol.*, 11(1):43–55.

Maarten Grootendorst. 2022. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.

William B Gudykunst, Stella Ting-Toomey, and Elizabeth Chua. 1988. *Culture and interpersonal communication.* Sage Publications, Inc.

S C Guntuku, M Li, L Tay, and L H Ungar. 2019a. Studying cultural differences in emoji usage across the east and the west. *Proceedings of the International*.

Sharath Chandra Guntuku, Mingyang Li, Louis Tay, and Lyle H Ungar. 2019b. Studying cultural differences in emoji usage across the east and the west. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, pages 226–235.

Sharath Chandra Guntuku, Daniel Preotiuc-Pietro, Johannes C Eichstaedt, and Lyle H Ungar. 2019c. What twitter profile and posted images reveal about depression and anxiety. In *Proceedings of the international AAAI conference on web and social media*, volume 13, pages 236–246.

Shreya Havaldar, Salvatore Giorgi, Sunny Rai, Thomas Talhelm, Sharath Chandra Guntuku, and Lyle Ungar. 2024. Building knowledge-guided lexica to model cultural variation. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 211–226, Mexico City, Mexico. Association for Computational Linguistics.

Geert Hofstede. 2011. Dimensionalizing cultures: The hofstede model in context. *Online readings in psychology and culture*, 2(1):8.

Kathleen L Hourihan, Scott H Fraundorf, and Aaron S Benjamin. 2017. The influences of valence and arousal on judgments of learning and on recall. *Mem. Cognit.*, 45(1):121–136.

Kokil Jaidka, Salvatore Giorgi, H Andrew Schwartz, Margaret L Kern, Lyle H Ungar, and Johannes C Eichstaedt. 2020. Estimating geographic subjective well-being from twitter: A comparison of dictionary and data-driven language methods. *Proc. Natl. Acad. Sci. U. S. A.*, 117(19):10165–10171.

Kokil Jaidka, Sharath Guntuku, and Lyle Ungar. 2018. Facebook versus twitter: Differences in Self-Disclosure and trait prediction. *ICWSM*, 12(1).

Lisa N Jefferies, Daniel Smilek, Eric Eich, and James T Enns. 2008. Emotional valence and arousal interact in attentional control. *Psychological science*, 19(3):290–295.

Margaret L Kern, Lea Waters, Alejandro Adler, and Mathew White. 2014. Assessing employee wellbeing in schools using a multifaceted approach: Associations with physical health, life satisfaction, and professional thriving. *Psychology*, 05(06):500–513.

Peter Kuppens, Francis Tuerlinckx, James A Russell, and Lisa Feldman Barrett. 2013. The relation between valence and arousal in subjective experience. *Psychol. Bull.*, 139(4):917–940.

Peter Kuppens, Francis Tuerlinckx, Michelle Yik, Peter Koval, Joachim Coosemans, Kevin J Zeng, and James A Russell. 2017a. The relation between valence and arousal in subjective experience varies with personality and culture. *Journal of personality*, 85(4):530–542.

Peter Kuppens, Francis Tuerlinckx, Michelle Yik, Peter Koval, Joachim Coosemans, Kevin J Zeng, and James A Russell. 2017b. The relation between valence and arousal in subjective experience varies with personality and culture. *J. Pers.*, 85(4):530–542.

P J Lang. 1994. The varieties of emotional experience: a meditation on James-Lange theory. *Psychol. Rev.*, 101(2):211–221.

Lung-Hao Lee, Jian-Hong Li, and Liang-Chih Yu. 2022. Chinese emobank: Building valence-arousal resources for dimensional sentiment analysis. *Transactions on Asian and Low-Resource Language Information Processing*, 21(4):1–18.

Lin Li, Ang Li, Bibo Hao, Zengda Guan, and Tingshao Zhu. 2014. Predicting active users' personality based on micro-blogging behaviors. *PloS one*, 9(1):e84997.

Mingyang Li, Sharath Guntuku, Vinit Jakhetiya, and Lyle Ungar. 2019. Exploring (dis-) similarities in emoji-emotion association on twitter and weibo. In *Companion proceedings of the 2019 world wide web conference*, pages 461–467.

Mingyang Li, Louis Hickman, Louis Tay, Lyle Ungar, and Sharath Chandra Guntuku. 2020. Studying politeness across cultures using english twitter and mandarin weibo. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2):1–15.

Zhongguo Li and Maosong Sun. 2009. Punctuation as implicit annotations for chinese word segmentation. *Comput. Linguist.*, 35(4):505–512.

Nangyeon Lim. 2016. Cultural differences in emotion: differences in emotional arousal level between the east and the west. *Integrative medicine research*, 5(2):105–109.

Youn-kyung Lim, Justin Donaldson, Heekyoung Jung, Breanne Kunz, David Royer, Shruti Ramalingam, Sindhia Thirumaran, and Erik Stolterman. 2008. Emotional experience and interaction design. *Affect and emotion in human-computer interaction: From theory to applications*, pages 116–129.

Xialing Lin, Kenneth A Lachlan, and Patric R Spence. 2016. Exploring extreme events on social media: A comparison of user reposting/retweeting behaviors on twitter and weibo.

Gloria Liou, Juhi Mittal, Neil KR Sehgal, Louis Tay, Lyle Ungar, and Sharath Chandra Guntuku. 2023. The online language of work-personal conflict. *Scientific Reports*, 13(1):21019.

Luo Lu, Robin Gilmour, and Shu-Fang Kao. 2001. Cultural values and happiness: An east-west dialogue. *The Journal of social psychology*, 141(4):477–493.

Wei Lü and Zhenhong Wang. 2012. Emotional expressivity, emotion regulation, and mood in college students: A cross-ethnic study. *Social Behavior and Personality: an international journal*, 40(2):319–330.

Marco Lui and Timothy Baldwin. 2012. langid.py: An off-the-shelf language identification tool. In *Proceedings of the ACL 2012 System Demonstrations*, pages 25–30.

Lin Ma. 2013. Electronic word-of-mouth on microblogs: A cross-cultural content analysis of twitter and weibo. *Intercultural Communication Studies*, 22(3).

Hazel R Markus and Shinobu Kitayama. 1991. Cultural variation in the self-concept. In *The self: Interdisciplinary approaches*, pages 18–48. Springer.

David Matsumoto. 1990. Cultural similarities and differences in display rules. *Motivation and emotion*, 14:195–214.

Andrew Kachites McCallum. 2002. Mallet: A machine learning for languagetoolkit. *http://mallet. cs. umass. edu*.

Yuri Miyamoto, Xiaoming Ma, and Amelia G Petermann. 2014. Cultural differences in hedonic emotion regulation after a negative event. *Emotion*, 14(4):804.

Saif Mohammad. 2018a. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Stroudsburg, PA, USA. Association for Computational Linguistics.

Saif Mohammad. 2018b. Word affect intensities. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Saif M Mohammad. 2016. Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In *Emotion measurement*, pages 201–237. Elsevier.

James W Neuliep. 2012. The relationship among intercultural communication apprehension, ethnocentrism, uncertainty reduction, and communication satisfaction during initial intercultural interaction: An extension of anxiety and uncertainty management (aum) theory. *Journal of Intercultural Communication Research*, 41(1):1–16.

Catherine J Norris, Jackie Gollan, Gary G Berntson, and John T Cacioppo. 2010. The current status of research on the structure of evaluative space. *Biological psychology*, 84(3):422–436.

Andrew Ortony, Gerald L Clore, and Allan Collins. 1990. *The Cognitive Structure of Emotions*. Cambridge University Press.

Wei Pan, Ren-jie Wang, Wan-qiang Dai, Ge Huang, Cheng Hu, Wu-lin Pan, and Shu-jie Liao. 2021. China public psychology analysis about covid-19 under considering sina weibo data. *Frontiers in Psychology*, 12:713597.

Daniel Preoţiuc-Pietro, H Andrew Schwartz, Gregory Park, Johannes Eichstaedt, Margaret Kern, Lyle Ungar, and Elisabeth Shulman. 2016. Modelling valence and arousal in facebook posts.

Paul R Rosenbaum and Donald B Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

James A Russell. 1980. A circumplex model of affect.

Bahar Salehi, Dirk Hovy, Eduard Hovy, and Anders Søgaard. 2017. Huntsville, hospitals, and hockey teams: Names can reveal your location.

Maarten Sap, Gregory Park, Johannes Eichstaedt, Margaret Kern, David Stillwell, Michal Kosinski, Lyle Ungar, and Hansen Andrew Schwartz. 2014. Developing age and gender predictive lexica over social media.

H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin E P Seligman, and Others. 2013a. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS One*, 8(9):e73791.

H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin EP Seligman, et al. 2013b. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one*, 8(9):e73791.

Gideon Schwarz. 1978. Estimating the dimension of a model. *aos*, 6(2):461–464.

Shula Sommers. 1984. Reported emotions and conventions of emotionality among college students. *Journal of Personality and Social Psychology*, 46(1):207.

Zihuang Tang, Peishan Wang, Xiaoyang Sui, and Yue Fan. 2019. Effects of psychological distance on the negative emotions of immoral events—a study based on weibo data. *Human Behavior and Emerging Technologies*, 1(3):208–215.

Michael A Tarrant, Michael J Manfredo, Peter B Bayley, and Richard Hess. 1993. Effects of recall bias and nonresponse bias on Self-Report estimates of angling participation. *N. Am. J. Fish. Manage.*, 13(2):217–222.

Xianyun Tian, Philip Batterham, Shuang Song, Xiaoxu Yao, and Guang Yu. 2018. Characterizing depression issues on sina weibo. *Int. J. Environ. Res. Public Health*, 15(4).

Jeanne L Tsai. 2007. Ideal affect: Cultural causes and behavioral consequences. *Perspectives on Psychological Science*, 2(3):242–259.

Jeanne L Tsai, Brian Knutson, and Helene H Fung. 2006a. Cultural variation in affect valuation. *J. Pers. Soc. Psychol.*, 90(2):288–307.

Jeanne L Tsai, Brian Knutson, and Helene H Fung. 2006b. Cultural variation in affect valuation. *Journal of personality and social psychology*, 90(2):288.

Yi Wenjia, Zhao Yanyan, Yuan Jianhua, Zhao Weixiang, and Qin Bing. 2023. Improving affective event classification with multi-perspective knowledge injection. In *Proceedings of the 22nd Chinese National Conference on Computational Linguistics*, pages 773–785, Harbin, China. Chinese Information Processing Society of China.

Eugene Winograd and Ulric Neisser. 2006. *Affect and Accuracy in Recall: Studies of 'Flashbulb' Memories*. Cambridge University Press.

Yuanyuan Xu, Yongju Yu, Yuanjun Xie, Li Peng, Botao Liu, Junrun Xie, Chen Bian, and Min Li. 2015. Positive affect promotes well-being and alleviates depression: The mediating effect of attentional bias. *Psychiatry Res.*, 228(3):482–487.

Michelle S M Yik, James A Russell, and Lisa Feldman Barrett. 1999. Structure of self-reported current affect: Integration and beyond. *J. Pers. Soc. Psychol.*, 77(3):600–619.

Wanru Zhang, Andrew Caines, Dimitrios Alikaniotis, and Paula Buttery. 2016. Predicting author age from weibo microblog posts. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 2990–2997.

Yuan Zhong, Nicholas Jing Yuan, Wen Zhong, Fuzheng Zhang, and Xing Xie. 2015. You are where you go.

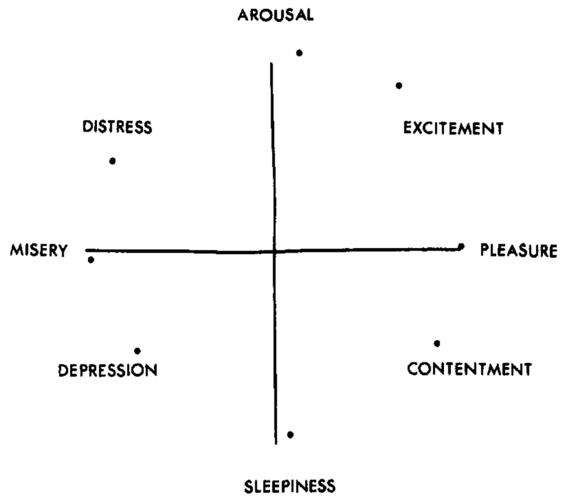


Figure 1. Eight affect concepts in a circular order.

Figure 6: Example of Russell's two-dimensional circumplex model. Captured from Russell 1980.

Appendix

A Details of Pearson r

In our paper, Pearson r correlation is independently calculated for post level valence and arousal scores, which gives each n-gram and topic a valence and arousal score. The Pearson r correlation coefficient is calculated with the OLS regression. Since the Coefficient in OLS regression is:

$$\beta = \frac{Cov(x, y)}{Var(x)} \tag{5}$$

And Pearson Correlation Coefficient is:

$$r = \frac{Cov(x, y)}{\sqrt{Var(x) \cdot Var(y)}} \tag{6}$$

So β can be represented by r with:

$$\beta = r \cdot \frac{SD(y)}{SD(x)} \tag{7}$$

So when normalized, $\beta = r$.

B Explanation of Conditional R^2

In Table 1, we use conditional R^2 to represent the variance explained by the entire model, including both fixed and random factors.

The method we used for finding functional relationships between valence and arousal is a follow-up analysis using the same method used in the previous work. Table 2 in Kuppens et al., 2013 listed BIC, PostP, and the best R^2 for different studies

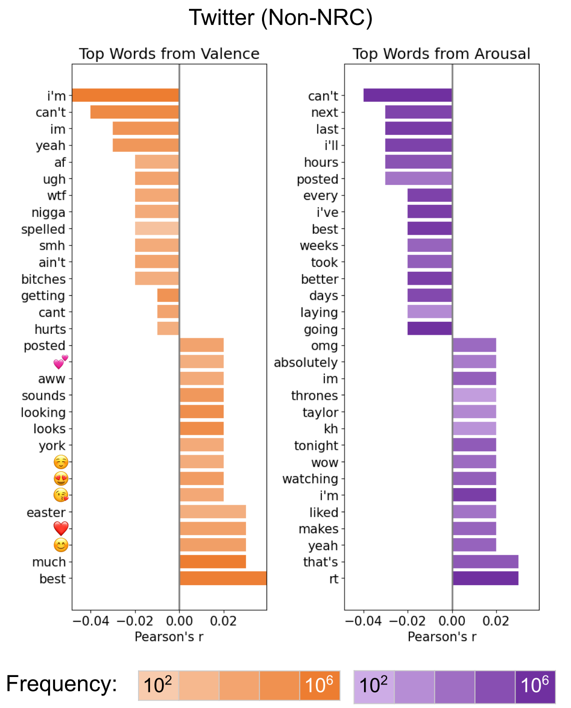


Figure 7: Words and phrases associated with valence and arousal on Twitter (Non-NRC) from the top 15 phrases for effect strength (Pearson r), colored by frequency. Statistically significant ($p < .05$, two-tailed t-test, Benjamini-Hochberg corrected).

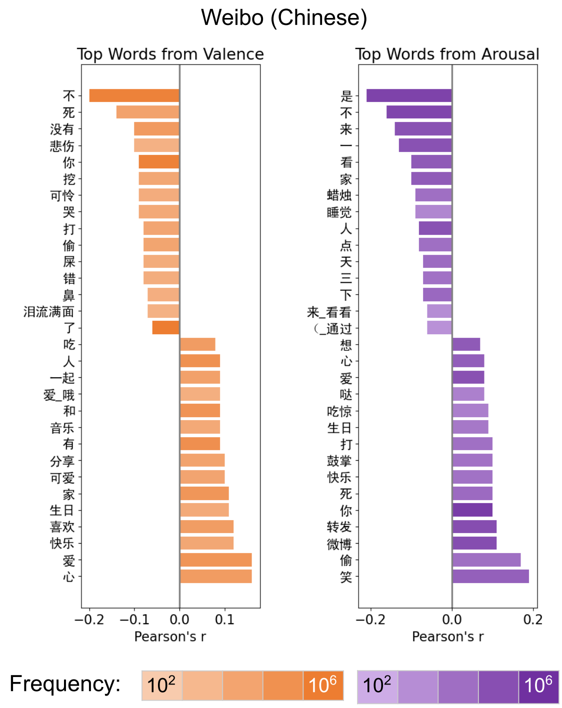


Figure 8: Words and phrases associated with valence and arousal on Weibo (Chinese) from the top 15 phrases for effect strength (Pearson r), colored by frequency. Statistically significant ($p < .05$, two-tailed t-test, Benjamini-Hochberg corrected).

and datasets, which are consistent with our result - relatively low R^2 scores. In our paper, we expand this work using social media data to see if a similar conclusion can hold across cultures with the functional relationship between valence and arousal.

As mentioned in (Kuppens, Tuerlinckx, Russell, and Barrett, 2013), low to moderate R^2 values of our functions are expected, not only because our analysis is based on noisy social media data, but also affective experiences of all combinations of valence and arousal can occur. For example, although less likely, the LDA result in our paper shows that positive valence low arousal states are represented by relaxing, and sleep.

Our goal in the analysis of functional relationships is not to train a model for predicting the arousal of a sentence using valence, but to explain the relationship between valence and arousal on average.

C Details of Valence-Arousal Relation Models

In this section, we show the full comparison between Model 1 to Model 6 in Figure 10. Among all, Model 6 fits the social media data best, shows negative offset and negative bias in the line of best fit. This is opposite to the previous findings where positive offsets were observed from lab-based measurements.

D Removal of NRC-VAD Lexicons

To give a multidimensional insight into culturally specific expressions of valence and arousal, we show Twitter part of Figure 3 without using words from NRC-VAD Lexicon in Figure 7. NRC-VAD contains 20,000 words, which covers most of the daily vocabulary. This figure, without lexicon words, consists of a lot of emojis, internet slang, and swear words.

E Pre-Translated Weibo Figures

All the figures and tables from Weibo are translated into English with Google Translate. Here, we show the figures with original Chinese text. Figure 8 shows the top 15 phrases for effect strength with valence and arousal. Figure 9 shows the top topics associated with valence and arousal.

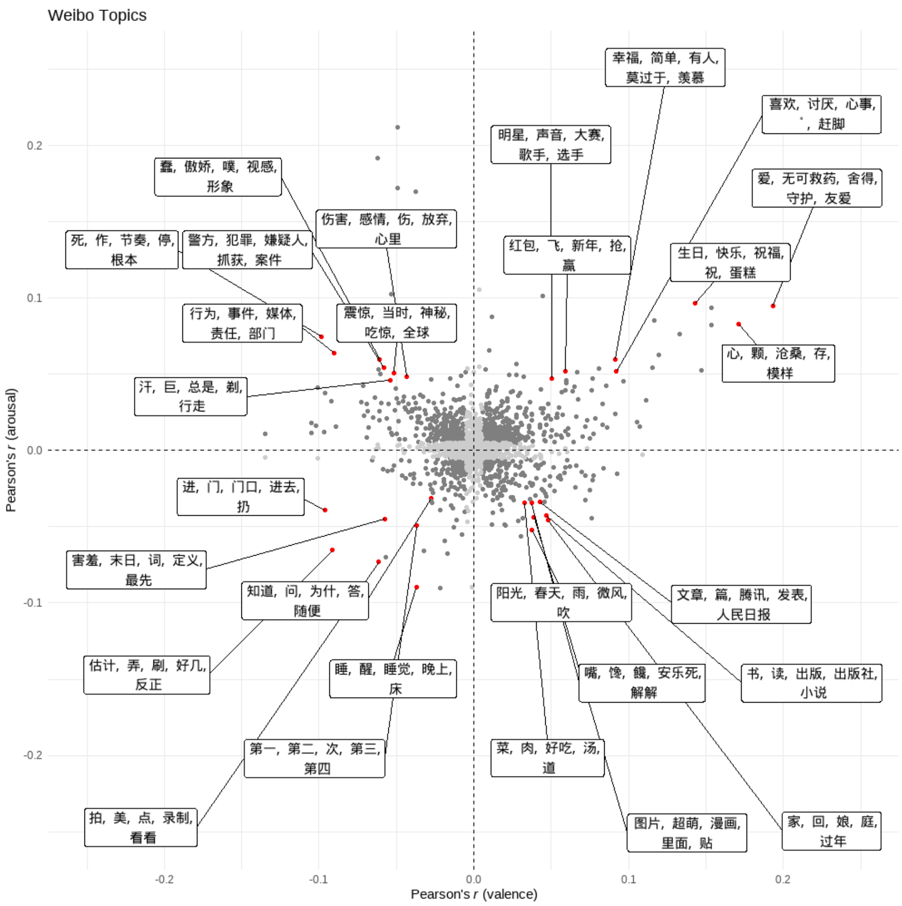
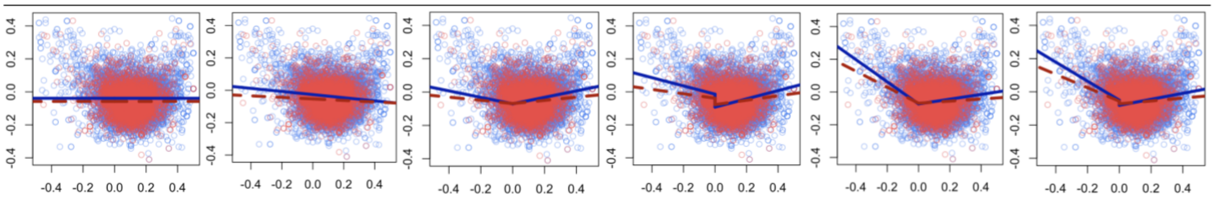


Figure 9: Topics associated with valence and arousal on Weibo (Chinese), sorted by effect size (Pearson r). Each point is a topic and statistically significant topics ($p < .05$, two-tailed t-test, Benjamini-Hochberg corrected) are shown in dark gray. The X-axis is the Pearson r with valence and the Y-axis with arousal. English translations of the top 5 words in each topic are shown.

Independence	Linear Relation	Symmetric V-shaped	Asymmetric V-shaped		
Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**



Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
$Arousal_m = \beta_0 + \varepsilon_m$	$Arousal_m = \beta_0 + \beta_1 Valence_m + \varepsilon_m$	$Arousal_m = \beta_0 + \beta_1 \lvert Valence_m \rvert + \varepsilon_m$	$Arousal_m = \beta_0 + \beta_1 \lvert Valence_m \rvert + \beta_2 I_m + \varepsilon_m$	$Arousal_m = \beta_0 + \beta_1 \lvert Valence_m \rvert + \beta_3 I_m \lvert Valence_m \rvert + \varepsilon_m$	$Arousal_m = \beta_0 + \beta_1 \lvert Valence_m \rvert + \beta_2 I_m + \beta_3 I_m \lvert Valence_m \rvert + \varepsilon_m$

Figure 10: Scatter plots of valence (x-axis) and arousal (y-axis) of Twitter (blue) and Weibo (red) posts. The lines of best fit for each model's function are appended to each plot (Twitter: solid line, Weibo: dashed line). Each model is tested with a within-person intercept and slope.