Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-025-05927-0 OPEN

Understanding gender and age differences in language use: cross-cultural insights from Weibo and Facebook

Dandan Pang[1,2,11]✉, Sharath Chandra Guntuku[3,4,5,11], Garrick Sherman[3], Tingting Liu[3,6], Sunny Rai[3,5], Young-Min Cho[4], Rian You[7], Arnold B. Bakker[8,9,10] & Lyle H. Ungar[3,4]

This study integrates social role theory and socioemotional selectivity theory to investigate the cultural universalities and differences in language use among male and female users across different age groups on Weibo and Facebook. By analyzing social media language, we aim to understand how gender and age influence linguistic patterns and reflect broader cultural norms and societal values. Aggregated language from Weibo and Facebook users ($N = 8728$ per platform; 665,377 and 742,418 posts, respectively) was analyzed by both a top-down closed-vocabulary (Linguistic Inquiry and Word Count) approach and a data-driven open-vocabulary (Differential Language Analysis) approach. Our findings support and extend social role theory, showing that female users on both platforms use more communal and relational language, while male users focus on agentic and task-oriented content. Cultural dimensions, such as collectivism and individualism, modulate the expression of social roles, with Weibo users adhering more closely to traditional gender norms compared to Facebook users. Our findings also validate and extend the socioemotional selectivity theory by demonstrating how cultural frameworks shape the specific ways aging individuals pursue emotional and social goals. For example, on both platforms, age-related language patterns reveal a U-shaped trend in positive emotions, with a decline in middle age and an increase in older adulthood, reflecting a universal shift toward emotionally meaningful goals. Additionally, older users on Weibo engage more in collectivistic themes, while their Facebook counterparts focus on personal well-being and social ties. These results highlight the complex interplay between culture, gender, and age in shaping language use on social media, providing valuable insights into the cultural and societal influences on communication.

[1] Institute for New Work, Business School, Bern University of Applied Sciences, Bern, Switzerland. [2] Department of Work and Organizational Psychology, Institute of Psychology, University of Bern, Bern, Switzerland. [3] Positive Psychology Center, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [4] Department of Computer and Information Science, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [5] Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, USA. [6] National Institutes of Health, National Institute on Drug Abuse, Baltimore, Maryland, USA. [7] Baystate Health, Springfield, Cambridge, Massachusetts, USA. [8] Center of Excellence for Positive Organizational Psychology, Erasmus University of Rotterdam, Rotterdam, Netherlands. [9] School of Management, Human Resource Management, University of Vaasa, Vaasa, Finland. [10] School of Economics and Business, University of Ljubljana, Ljubljana, Slovenia. [11] These authors contributed equally: Dandan Pang, Sharath Chandra Guntuku.
✉email: dandan.pang@bfh.ch

Introduction

Language serves as a vital medium for expressing thoughts, emotions, and identities shaped by both personal experiences and the surrounding social and cultural contexts (Joshi et al., 2020). As digital communication becomes increasingly central to everyday life, social media platforms like Weibo and Facebook have emerged as powerful arenas where these expressions are made visible at scale (Cho et al., 2025; Cui et al., 2022; Guntuku et al., 2019). More than just channels of interpersonal interaction, social media functions as a cultural mirror, reflecting how individuals perform identity within culturally embedded norms and social expectations (Papacharissi, 2011).

Importantly, these platforms do not exist in a cultural vacuum. As Burger, Thornborrow, and Fitzgerald (2017) note, social media provide "rich and varied environments for interactional encounters" (p. 25), while simultaneously introducing new rules, symbols, and communicative constraints that mirror—and sometimes shape—offline culture. In China, for example, the internet is shaped by state regulation and collective values around community and personhood, while Western platforms often reflect more individualistic and expressive norms (Ju et al., 2019). Such cultural orientations are also evident in interpersonal behavior: while users from Asian contexts often emphasize relationship-building before task completion, Western users tend to prioritize the reverse—patterns that extend into online spaces (Sandel et al., 2019).

Moreover, unlike controlled laboratory settings or self-report surveys, social media provide naturally occurring, real-time data that captures spontaneous language use. This makes it a particularly valuable resource for examining how psychological characteristics—such as gender and age—interact with cultural frameworks to shape communication. While substantial work has established that language reflects gender roles and age-related changes (Garimella et al., 2016; Kern et al., 2014; Loveys et al., 2018; Schwartz et al., 2013c), less is known about how these patterns manifest across different cultural settings, particularly between collectivistic societies like China and individualistic societies like the United States (Garimella et al., 2016; Loveys et al., 2018).

To explore this question, we focus on two culturally distinct platforms: Weibo, the dominant microblogging site in China, and Facebook, a widely used social network in the United States. These platforms were selected not just for their popularity during the time of data collection (i.e., 2014), but because they offer culturally embedded spaces that align with the broader societal norms of their respective countries. While we recognize that platform-specific features (e.g., content censorship, self-disclosure norms) may influence communication styles, we treat these factors as part of the broader cultural ecosystems rather than as methodological confounds. Rather than assuming strict equivalency between the two platforms, we leverage their differences to investigate how cultural values are reflected—and potentially reinforced—through everyday digital language use.

This study integrates social role theory (SRT; Martin and Ruble, 2010; Wood and Eagly, 2015) and socioemotional selectivity theory (SST; Carstensen et al., 2003; Carstensen et al., 1999) as core framework to examine the cultural universalities and differences in gender- and age-related language use, focusing on Chinese (Weibo) and United States (Facebook) cohorts. We aim to address two main research questions: (1) How do gender and age manifest in language patterns on social media? and (2) How do cultural contexts shape these manifestations differently on Weibo and Facebook? By exploring these questions, we extend prior research to identify whether linguistic trends are universal or culturally specific.

Theoretical framework

Cultural differences between China and the United States. Culture has been defined as a constellation of loosely organized values, practices, and norms shared by an interconnected group of people in a given nation (Chiu et al., 2010). A foundational model for comparing cultures was developed by Geert Hofstede (1980), whose research with IBM employees across 40 countries identified four basic dimensions of cultural values, most notably collectivism versus individualism (Hofstede, 2001; Kim et al., 1994). This dimension remains central to understanding cultural differences across nations, particularly between China and the United States. In Chinese culture, collectivism prevails, emphasizing group harmony and prioritizing community welfare over individual pursuits. Rooted in Confucian, Buddhist, and Taoist values, this cultural framework fosters modesty and self-effacement (Chiu et al., 2010; Fu et al., 2007). Conversely, American culture exemplifies individualism, which values personal autonomy and pride in self-promotion (Wang and Leichtman, 2000).

The contrast between collectivism and individualism extends into the construct of self-construals: individuals in Eastern cultures typically adopt interdependent self-construals, emphasizing relational identities and interconnectedness, as the core unit of society is the group. In addition, individuals must adjust to the group so that society's harmony is maintained (Oyserman and Lee, 2008). In contrast, those in Western cultures tend to exhibit independent self-construals, valuing uniqueness and personal autonomy (Markus and Kitayama, 1991). This distinction also aligns with the theory of cultural tightness-looseness, where Eastern cultures are generally *tight*, with stronger social norms and lower tolerance for deviance, whereas Western cultures display greater *looseness*, characterized by more flexible social norms (Gelfand et al., 2011; Uz, 2015).

Emotion expression across cultures. Cultural norms shape not only how people experience emotions but also how they express them linguistically. One influential framework that captures these differences is the two-dimensional model of emotion (Russell, 1980), which distinguishes emotional states along two axes: valence (positive to negative) and arousal (high to low intensity). While much emotion research focuses on valence (e.g., happiness vs. sadness), Lim (2016) highlights that cultural differences are especially pronounced in terms of arousal. In East Asian cultures, influenced by Confucian values and interdependent self-construals, individuals tend to prefer low-arousal positive states (e.g., calmness, contentment) and suppress high-arousal negative emotions (e.g., anger) to maintain social harmony (Lim, 2016; Sims et al., 2015; Tsai et al., 2006). In contrast, Western cultures, influenced by autonomy and independent self-construals, tend to value and promote high-arousal positive states (e.g., excitement, enthusiasm) and tolerate more direct expressions of high-arousal negative emotions, particularly when asserting personal boundaries or expressing dissatisfaction (Lim, 2016; Sims et al., 2015; Tsai et al., 2006).

These differing affective ideals are reflected in language. In collectivistic societies, emotional expression often aims to preserve group cohesion; strong emotions may be softened or blended to avoid discomfort or disruption. By contrast, individualistic cultures encourage emotional amplification. These culturally rooted tendencies in emotion regulation and expression create distinct linguistic patterns, which are observable on social media platforms like Weibo and Facebook and offer insights into how cultural norms interact with gender and age in shaping language use.

Gender differences in language use across cultures. Gender refers to the socially constructed roles, behaviors, activities, and attributes a given society considers appropriate for men and women (World Health Organization, 2014). Gender differences, as explained by social role theory (Martin and Ruble, 2010; Wood and Eagly, 2015), are primarily driven by the division of labor derived from male and female biology, such as reproductive functions and physical attributes. This division of labor is then reinforced and legitimized through gender stereotypes, which makes it appear natural and inevitable. These stereotypes and roles are universal, but they are also significantly influenced by cultural and societal values (Wood and Eagly, 2015).

Both East Asian and Western cultures share major gender norms and roles: women are generally expected to display femininity/communion—characterized by gentleness, empathy, emotional sensitivity, and expressiveness—while men are expected to show masculinity/agency—characterized by assertiveness, independence, and courage (Louie, 2002). Additionally, women are more likely to take on relational roles (e.g., caregiving), while men are more likely to take on task-oriented roles (e.g., earning a living), both at home and at work (Hsu et al., 2021; Martin and Ruble, 2010; Wood and Eagly, 2002, 2015). The concepts of agency and communion represent the fundamental modalities of human nature. A recent meta-analysis by Eagly et al. (2020) found that the past 73 years have produced an accentuated stereotype of women as the more communal sex, with men retaining their agency advantage.

However, much research on gender has generalized observed differences without addressing their cultural contexts, leading to gaps in understanding how these gender differences might vary or persist outside specific research settings. Notably, cultural factors such as the tightness-looseness dimension (Gelfand et al., 2011; Uz, 2015) suggest that East Asian societies, such as China, tend to reinforce traditional gender norms more strictly than Western societies (Liu and Iwamoto, 2006). This aligns with data from the Global Gender Gap Index 2020, which highlights a wider gender gap in China compared to the United States, indicating more pronounced gender biases in work and life contexts in China (World Economic Forum, 2020).

In summary, while gender roles and stereotypes are universal, cultural and societal values may shape their expression. Examining these differences through social media language offers valuable insights into the generality and variability of gender norms across cultures.

Age-related changes in language use across cultures. Age-related changes, much like gender differences, are significant in shaping human behavior and experiences. Socioemotional selectivity theory (Carstensen et al., 2003; Carstensen et al., 1999) provides a comprehensive understanding of how aging influences affective development of humans. SST suggests that as individuals age, their perception of time shifts, leading them to prioritize emotionally and socially meaningful goals.

Contrary to the common perception of aging as merely a period of decline, it is also marked by positive growth, including emotional development and personality maturation. Studies have revealed a U-shaped trajectory in positive emotions across the lifespan: positive emotions peak during young adulthood, dip during middle age, and rise again in later years, with an upswing often beginning in the 50s (Blanchflower and Graham, 2022). According to socioemotional selectivity theory, this trend arises because individuals, perceiving their remaining time as limited, prioritize experiences that foster emotional satisfaction and strengthen meaningful social bonds. These emotional shifts are accompanied by personality development, as aging individuals tend to become more responsible, caring, self-directed, and cooperative (Josefsson et al., 2013). In older adulthood, the motivation for generativity—the desire to guide and support the next generation—becomes particularly significant (McAdams et al., 1993; Villar, 2012).

This developmental process highlights how aging is not solely about loss but also about selective optimization and maturity, where individuals refine their focus on what matters most and take on social responsibility. However, while age-related affective and social shifts are universal, cultural contexts play a critical role in shaping how these changes are expressed, particularly in language. For instance, older individuals in East Asian cultures tend to display stronger collectivistic values, emphasizing self-control, social harmony, and group cohesion, while their Western counterparts often prioritize individual well-being and personal autonomy (Labouvie-Vief et al., 2000).

We believe that these age-related shifts will also be reflected in social media language, providing valuable insights into the interplay between psychological development and cultural influences. This approach is particularly valuable because it leverages social media language as a real-time, naturalistic, and ecologically valid form of communication. Traditional psychological research often relies on controlled environments or self-reported data, which can limit the authenticity and representativeness of the findings. Social media, on the other hand, offers a vast and diverse dataset of spontaneous language use that mirrors individuals' everyday interactions, providing a more accurate reflection of how people express emotions, social roles, and identity in natural contexts.

Hypotheses development

Building on these frameworks, we hypothesize that cultural contexts will modulate the gendered and age-related linguistic patterns predicted by social role theory and socioemotional selectivity theory. Specifically, we propose the following hypotheses:

Hypothesis 1: Cultural norms and emotional expression (Cultural Norms).

On Weibo, emotional expression is expected to reflect greater moderation and balance, characterized by a mix of positive and negative emotions, consistent with collectivistic cultural values and the goal of maintaining social harmony. In contrast, emotional expression on Facebook is expected to exhibit stronger emotional arousal, including both high-arousal positive emotions (e.g., excitement, enthusiasm) and high-arousal negative emotions (e.g., anger), in line with individualistic cultural norms that promote self-expression and emotional intensity.

Hypothesis 2: Gendered language differences (SRT)

Across both Weibo and Facebook, female users are expected to use more communal and relational language, such as words related to affiliation, emotion, and social processes. In contrast, male users are expected to use more agentic and task-oriented language, such as words related to achievement, objects, and power.

Hypothesis 3: Cultural reinforcement of gender norms (SRT x Cultural Norms)

Gendered language differences are expected to be more pronounced on Weibo than on Facebook, due to stronger cultural enforcement of traditional roles in collectivistic and tight cultures like China, compared to individualistic and loose cultures like the U.S.

Hypothesis 4: Age-related shifts in positivity (SST)

Across both platforms, we expect a U-shaped relationship between age and the use of positive emotional language, with a

decline during midlife and an increase in older adulthood, as older users prioritize emotionally meaningful content.

Hypothesis 5: Age-related shifts in social focus (SST x Cultural Norms)

With increasing age, users on both platforms are expected to shift from self-focused to socially meaningful language. However, older Weibo users may emphasize collectivistic values (e.g., group identity, duty), while older Facebook users may prioritize individual well-being and social ties, reflecting cultural norms of interdependence vs. independence.

This study contributes to the literature in three key ways. First, it provides cross-cultural insights by identifying both universal and culturally specific patterns in gender and age-related language, enhancing our understanding of the dynamic interplay between culture and psychology. Second, it offers real-world validation by demonstrating how social media, as a real-time and naturalistic medium, can extend theoretical frameworks and provide empirical support. Third, it highlights practical applications by underscoring the role of language in reflecting and reinforcing societal norms, with implications for cross-cultural communication, digital platform design, and artificial intelligence. By addressing these hypotheses, this study advances our knowledge at the intersection of culture, psychology, and digital communication, revealing how societal values are embedded in everyday language use across diverse cultural contexts.

Methods

Data. Our materials are status updates posted on Weibo and Facebook. Weibo posts were collected in 2014 using a breadth-first search (BFS) web crawling strategy from prior work (Cho et al., 2025; Cui et al., 2022; Guntuku et al., 2019). This produced a large dataset of over 29 million posts from 859,054 users in 2014. Similar to prior work (Guntuku et al., 2019), traditional Chinese characters were converted to Simplified Chinese using the *hanziconv* Python package to conform to the LIWC dictionary. We also remove any direct re-posts (indicated by '@USERNAME//'). From this dataset, we selected a subset of users who (a) had self-reported gender and birth year data[1], and (b) wrote at least 500 words in total across their posts. This filtering resulted in a final Weibo sample of 8728 users with 665,377 posts. Facebook data were obtained from the MyPersonality project, which collected user-generated content between 2010 and 2011 (Stillwell and Kosinski, 2015). In this dataset, users voluntarily shared their status updates and demographic information. We selected users who (a) were English-speaking, (b) were under the age of 65, and (c) wrote at least 500 words in their status updates. While specific geographic location was not recorded in the MyPersonality dataset, previous research (e.g., Schwartz et al., 2013a) has found that this sample is primarily representative of English-speaking users, predominantly based in the United States. To ensure demographic comparability across platforms, we drew a stratified subsample of 8728 Facebook users to exactly match the Weibo sample's distribution by gender and age group. This approach allowed for cross-platform comparison while controlling for demographic composition. This matching procedure is reflected in Table 1, which provides an overview of participant demographics by gender and age group on Weibo and Facebook. We used 500 words as a threshold based on prior works that looked at power analyses of predictive performance when working with social media data to analyze individuals' personal characteristics (Jaidka et al., 2018). Table 1 provides an overview of participant demographics by gender and age group on Weibo and Facebook. The users in this study did not receive any compensation for their contributions.

Linguistic analyses

Language pre-processing. We used *Langid* (Lui et al., 2012) to detect the language and filter posts not written in Mandarin on Weibo and English on Facebook, thereby eliminating the confounds of bilingualism (Fishman, 1980). We then split ('tokenized') the Weibo and Facebook posts into tokens, punctuation, and emoticons. Weibo posts were segmented into tokens using *THULAC* (Sun et al., 2016) while Facebook posts were segmented using DLATK/ *happierfuntokenizing* (Potts, 2017). All tokens used by less than 1% of users were removed from the analysis to eliminate uncommonly used tokens (outliers). Language follows a Zipfian distribution, in which the vast majority of tokens are used very infrequently; by dropping these tokens, we ensure that the language identified in our analysis is common enough to generalize to out-of-sample instances.

Closed vocabulary. First, we subextracted 71 categories in Mandarin (Huang et al., 2012) and 73 subcategories in English provided by Linguistic Inquiry and Word Count (LIWC; Pennebaker et al., 2015). Sample dictionaries included psychological (e.g., positive emotion), social (e.g., family), and syntactic categories (e.g., pronouns). The relative frequency of each dictionary—the total number of times a word written by the user matches a word in a given dictionary, divided by the user's total number of words—was also extracted. To test our theory-driven hypotheses, we mapped specific LIWC categories to relevant psychological constructs. For example, in Hypothesis 3, communal and relational language was operationalized using LIWC categories such as affective processes (i.e., positive emotion and negative emotion), social processes (e.g., family and friends), and personal pronouns to reflect a more emotionally engaged and interpersonal communication style. Conversely, agentic and task-oriented language was captured through personal concerns categories such as work and money, which align with goal-directed and achievement-oriented communication patterns. In Hypothesis 5, Self-focused language was measured using first-person singular pronouns (e.g., I and me), while socially meaningful language was captured through social processes (e.g., family and friends) and first-person plural pronouns (e.g., we). These operationalizations are grounded in prior research linking these LIWC categories to gendered communication styles and social role theory (e.g., Newman et al., 2008; Pennebaker et al., 2015).

Open vocabulary. Second, we performed differential language analysis (DLA, Schwartz et al., 2013b) to identify the most distinguishing language features for our outcomes. We extracted phrases consisting of one word, two or three consecutive words (called 1–3 grams) in the present study, for details of the methodology see (Kern et al., 2014; Schwartz et al., 2013a; Schwartz et al., 2013b). We kept only those 2- and 3-word phrases with high pointwise mutual information (PMI = 5; Church and Hanks, 2002), a ratio of the probability of observing the constituent words independently.

Third, we used the MALLET implementation of latent Dirichlet allocation (LDA; Blei et al., 2003) to identify latent data-driven word clusters. Although newer topic modeling techniques exist, research has shown that LDA remains effective while also maintaining interpretability (Dixon et al., 2022). We obtained 2000 topics for each Weibo and Facebook sample and for each user, we extracted the relative use of these 2000 topics.

Statistical analyses. Gender was coded as a binary variable based on self-report (0 = male, 1 = female), and age at the time of data collection was treated as a continuous variable. We examined the

Table 1 Participant demographics by gender and age group on Weibo and Facebook.

Age group	Weibo						Facebook					
	Female		Male		Total		Female		Male		Total	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
13–18	577	6.61	187	2.14	764	8.75	577	6.61	187	2.14	764	8.75
19–22	1800	20.62	675	7.73	2475	28.36	1800	20.62	675	7.73	2475	28.36
23–29	2794	32.01	1382	15.83	4176	47.85	2794	32.01	1382	15.83	4176	47.85
30–44	575	6.59	619	7.09	1194	13.68	575	6.59	619	7.09	1194	13.68
45–64	30	0.34	89	1.02	119	1.36	30	0.34	89	1.02	119	1.36
Total	5776	66.18	2952	33.82	8728	100.00	5776	66.18	2952	33.82	8728	100.00

Facebook users were selected using stratified sampling to exactly match the gender and age group distribution of Weibo users.

relationship between each demographic variable (gender or age), controlling for the other variable (i.e., when correlating with gender, we controlled for age, and vice versa), and three sets of linguistic features: (a) closed vocabulary LIWC, (b) open vocabulary 1–3 grams, and (c) open vocabulary topics. Analyses were conducted separately for Mandarin (Weibo) and English (Facebook) data, and we report the most strongly[2] associated language features for each.

Although the reported gender available for our data was dichotomous, the resulting Pearson's coefficient (r) is mathematically equivalent to a point-biserial correlation, a widely accepted effect size measure for binary comparisons (Cohen, 1988). We utilized Benjamini-Hochberg p-correction to correct for multiple comparisons and used $p < 0.005$ to indicate meaningful correlations. One of our hypotheses posits that aging affects specific language categories, such as positive emotions, in a curvilinear (U-shaped) pattern. To investigate this, we plotted age-related changes in positive emotions, using a figure to visualize these patterns and assess the U-shaped trend.

For Facebook users, we ranked LIWC subcategories based on the magnitude of r and selected the top 15 positive and negative correlations that were statistically significant based on Bonferroni-corrected p-values ($p < 0.005$). These subcategories were then grouped under their respective supercategories (based on the English LIWC hierarchy) for clarity. In some cases, the order of presentation reflects grouping rather than numerical rank. For Chinese LIWC, which lacks a direct matched category hierarchy, we aligned categories according to their English equivalents where possible and marked unmatched ones as unique (see Table 2 note).

Results

Gender

Closed vocabulary. Our correlational analysis showed acceptable correlations between gender and the LIWC lexicon categories for both Weibo and Facebook users (see Table 2). Among the top-correlated LIWC categories, biological process including body (e.g., neck), ingestion (e.g., eat, cook), and sexual (e.g. sex, nude, libido) generated the strongest correlations with female gender on Weibo, which indicates that females pay attention to and talk about physical appearance. Another group of top-correlated LIWC topics with female gender referred to self-representation, including both second and first personal pronouns (e.g., you, I, and myself), suggesting a more personal and relational style of communication. Our results also showed that females tend to use adverbs (e.g., once, gradually), auxiliary (e.g., not necessary, possible), and other function and grammar words in their Weibo postings, showing a degree of humility and indirectness. The top-correlated LIWC categories that are associated with female gender on Facebook were affective processes such as positive emotion, social processes such as family, personal concerns such as home, pointing to a willingness to share their emotional, relational and domestic topics in their life.

In contrast, the top-correlated LIWC categories with males for both Weibo and Facebook were personal concerns such as work (e.g., factory and interview) and money (e.g., wealth and annual salary), which indicates that males talk more about jobs and money-related activities in their postings than females. Males on both platforms also use a more quantitative way of description in their postings, including categories of articles (e.g., "this" and "that" on Weibo; "the" on Facebook), space (e.g., inside and down), and numbers. Males on both platforms shared more posts with swear words and words referencing death. In addition, different from Weibo findings, male gender correlated more strongly with anger on Facebook, pointing to a more direct and assertive communication style that is culturally permissible for males in the West.

These results from the closed vocabulary support *Hypothesis 2*, showing that female users on both Weibo and Facebook display a higher frequency of communal content, while male users on both platforms exhibit agency-related themes. The findings also confirm *Hypothesis 3*, as female Weibo users focus more on topics like body and self-representation, reflecting culturally expected gender roles around physical appearance. This highlights a stronger emphasis on traditional gender norms on Weibo compared to Facebook.

Open vocabulary. The 100 most distinctive words and phrases, as well as the LDA topics for each gender on Weibo (Fig. 1) and Facebook (Fig. 2), are visualized (see Supplementary Fig. S1 for a visualization in the original Chinese language).

In general, words and phrases associated with female posts on Weibo expressed mixed emotional feelings, which include both positive emotions (e.g., good, happy) and negative emotions (e.g., cry, shy). Consistent with the LIWC results, words that emphasize bodily looks and appearance (e.g., good-looking, beautiful), as well as personal pronouns (e.g., I, you) stand out. Additionally, LDA topics associated with women's language on Weibo also centered around mixed emotions (e.g., happy, shy) and appearance (e.g., ugly, face), with topics also covering celebrities such as 'Xiaopeng Li' and 'Ooli, a father-daughter duo that gained popularity from a famous TV show called 'The Return of Superman'. Similarly, female users' Facebook language was also associated with positive emotions (e.g., happy, wonderful, excited), celebratory expressions (e.g., birthday, yay), personal relationships (e.g., friends, mom), and appearances (e.g., beautiful, hair).

Within the female word clouds of Weibo, a notable pattern emerges with the prevalence of adjectives like cute and small. These terms are closely associated with the cultural construct of

Table 2 The top 15 most positively and negatively correlated linguistic inquiry and word count (LIWC) categories with gender on Weibo and Facebook.

Weibo						Facebook					
Super-category	Subcategories	Representative words	*r*	95% CI		Super-category	Subcategories	Representative words	*r*	95% CI	
				L	U					L	U
Positively correlated											
Biological processes	Body	脖子, 皮肤, 肠胃 (neck, skin, stomach)	0.245	0.225	0.265	Affective processes	Positive emotion	love, nice, sweet	0.225	0.205	0.245
	Ingestion	消化, 吃, 煮菜 (digest, eat, cook)	0.177	0.156	0.197		Sadness	crying, grief, sad	0.065	0.044	0.086
	Sexual	上床, 性欲, 裸体 (sex, libido, nudity)	0.151	0.131	0.172	Social processes	Family	daughter, dad, aunt	0.216	0.196	0.236
Personal pronouns	2nd person singular	你, 妳, 您 (you)	0.215	0.195	0.235		Female references	girl, her, mom	0.201	0.180	0.221
	1st person singular	本人, 自己, 我 (me, myself, I)	0.170	0.149	0.190		Friends	buddy, neighbor	0.083	0.062	0.104
Function words	Adverbs	曾经, 渐渐, 那么 (once, gradually, then)	0.232	0.212	0.251	Personal concerns	Home	kitchen, landlord	0.178	0.157	0.198
	Auxiliary verbs	不必, 可能, 应该 (not necessary, possible, should)	0.180	0.159	0.200	Informal language	Netspeak	btw, lol, thx	0.155	0.135	0.176
Chinese grammar	Multiple function[c]	的, 有, 是 (of, have, is)	0.128	0.107	0.149		Nonfluencies	er, hm, umm	0.066	0.045	0.087
	Interjunction[c]	呢, 吗, 吧[a] (*ne, ma, ba*)	0.221	0.201	0.241	Personal pronouns	1st person singular	I, me, mine	0.131	0.110	0.151
Perceptual processes	See	外貌, 闪亮, 绿色 (appearance, shiny, green)	0.135	0.114	0.156		3rd person singular	she, her, him	0.089	0.068	0.110
Social processes	Family	公婆, 兄弟, 孙女 (in-laws, brothers, granddaughters)	0.137	0.116	0.158	Drives	Affiliation	ally, friend, social	0.120	0.099	0.141
Cognitive processes	Discrepancy	不足, 纳闷, 期待 (lack, wonder, expect)	0.171	0.151	0.192	Biological processes	Health	clinic, flu, pill	0.082	0.061	0.103
	Inclusive	包括, 附近, 添加 (include, near, add)	0.157	0.137	0.178		Ingestion	dish, eat, pizza	0.075	0.054	0.096
	Certainty	不容置疑, 必然, 保证 (undoubtedly, inevitable, guaranteed)	0.139	0.118	0.160	Time orientations	Time	end, until, season	0.077	0.056	0.098
Other grammar	Common verbs	分享, 开车, 听 (share, drive, listen)	0.125	0.104	0.145		Future focus	may, will, soon	0.068	0.047	0.089
Negatively correlated											
Personal concerns	Work	工厂, 面试, 薪水(factory, interview, salary)	−0.204	−0.224	−0.183	Personal concerns	Death	bury, coffin, kill	−0.170	−0.190	−0.149
	Money	富有, 年薪, 折扣(wealth, annual salary, discount)	−0.146	−0.167	−0.126		Work	job, majors, xerox	−0.076	−0.097	−0.055
	Death	亡故, 自杀, 遗嘱 (death, suicide, will)	−0.037	−0.058	−0.016		Money	audit, cash, owe	−0.072	−0.093	−0.051
	Religion	上帝, 慈悲, 信仰 (god, mercy, faith)	−0.032	−0.053	−0.011	Affective processes	Anger	hate, kill, annoyed	−0.155	−0.175	−0.135
Chinese grammar	Specific article[c]	本, 该, 每 (this, that, every)	−0.110	−0.131	−0.089	Function words	Articles	a, an, the	−0.152	−0.173	−0.132
	Quantity unit[c]	条, 头, 枝 (strips, heads, branches)	−0.079	−0.099	−0.058	Informal language	Swear words	fuck, damn, shit	−0.146	−0.166	−0.125

Table 2 (continued)

Weibo						Facebook					
Super-category	**Subcategories**	**Representative words**	***r***	**95% CI**		**Super-category**	**Subcategories**	**Representative words**	***r***	**95% CI**	
				L	**U**					**L**	**U**
Relativity	Space	里面, 街道, 台上 (inside, street, on stage)	−0.106	−0.127	−0.085	Biological processes	Sexual	horny, love, incest	−0.128	−0.148	−0.107
	Postposition[c]	之中, 以上, 为止(among, above, till)	−0.100	−0.120	−0.079	Drives	Power	superior, bully	−0.124	−0.145	−0.103
	Time	期间, 过去, 秋天(period, past, fall)	−0.016	−0.037	0.005		Achievement	win, success, better	−0.088	−0.109	−0.067
Tense marker[c]	Present tense marker[c]	目前, 现今, 平时(currently, now, usually)	−0.102	−0.123	−0.081		Risk	danger, doubt	−0.075	−0.096	−0.054
	Past tense marker[c]	去年, 刚才, 昔日(last year, just now, in the past)	−0.039	−0.060	−0.018	Other Grammar	Numbers	second, thousand	−0.099	−0.120	−0.078
Function words	Prepositions	从, 依照, 把 (from, follow, refer to give/take something)	−0.063	−0.084	−0.042		Compare	greater, best, after	−0.074	−0.094	−0.053
Informal language	Swear words	去你的, 智障, 机车[b] (fuck you, mentally retarded, annoying)	−0.040	−0.061	−0.019		Interrogatives	how, when, what	−0.067	−0.088	−0.046
Other grammar	Numbers	一, 八百, 两万 (one, eight hundred, twenty thousand)	−0.035	−0.056	−0.014	Relativity	Space	down, in, thin	−0.080	−0.101	−0.060
Drives	Achieve	擅长, 尽责, 高手 (be good at, responsible, master)	−0.022	−0.043	−0.001	Cognitive processes	Tentative	maybe, perhaps	−0.062	−0.083	−0.041

LIWC category correlations. Gender (0 = male, 1 = female); 95% CI = 95% confidence interval, L = Lower, U = Upper; *r* = standardized linear regression coefficients.
All correlations in the table were significant ($p < 0.005$; BH *p*-corrected). For interpretability, subcategories are grouped under their LIWC supercategories based on the English LIWC structure.
[a]Markers for the perfective aspect in standard Chinese grammar expressing completion.
[b]A more typical phrase used in Taiwan than in Mainland China.
[c]Category unique to Chinese LIWC, with no English equivalent.

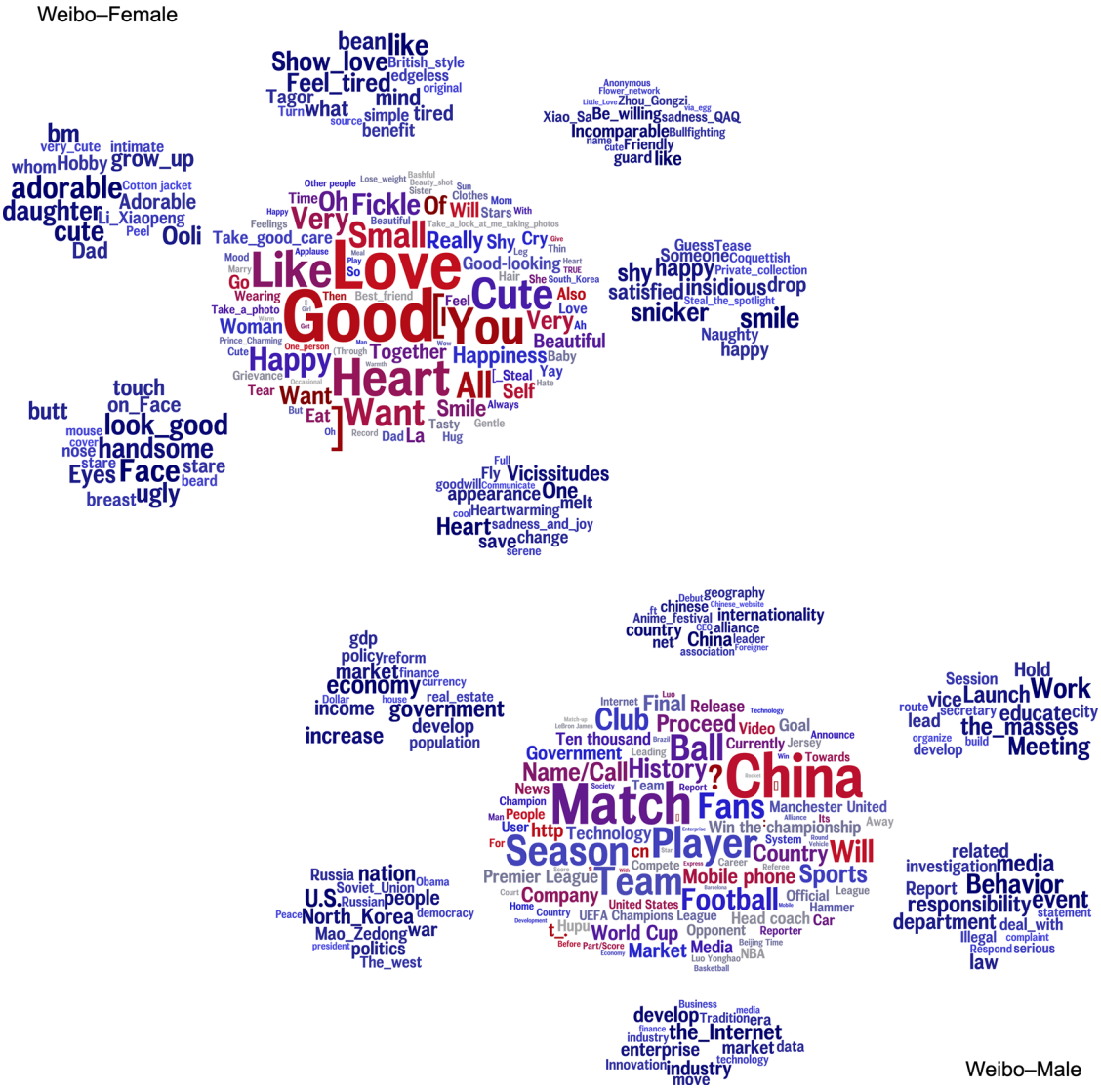

Fig. 1 Words and phrases for female and male users on Weibo. The 100 most distinctive words and phrases, as well as the latent Dirichlet allocation topics for each gender on Weibo. Results of female users are on the top left, and results of male users are on the bottom right. Correlations were adjusted for age and statistically significant ($p < 0.05$, two-tailed *t*-test, Benjamini–Hochberg corrected).

'kawaii/可爱' in East Asian contexts, which valorizes youthfulness and charm—attributes traditionally admired in female behavior as part of a broader societal preference for cuteness. In contrast, the word cloud generated from female users on Facebook shows a starkly different emotional landscape. Words such as excited, combined with the frequent use of exclamation marks, illustrate a mode of expression that conveys high emotional arousal.

Words and phrases associated with male posts showed very different patterns. Male users on Weibo leaned towards topics of national and political significance, sports fandom, and socio-economic matters. Words such as Mao Zedong, China, United States, political, and democracy signal a strong engagement with national identity and political discourse. In addition to this civic-minded conversation, there is significant talk about sports, indicating a shared enthusiasm for athletic competitions and events, as highlighted by terms like game, player, team, soccer, and World Cup. The focus on jobs, education, society, economy, and legal issues underlines a collective concern for the country's progress and ethical stance.

Conversely, male language on Facebook, while also encompassing sports-related enthusiasm and political discourse, diverges significantly in tone and content. The use of terms such as football, team, and game shows a common ground with Weibo in terms of sports interest. However, the presence of profanity and intense expressions such as hell and fucking in their language indicates a more informal and emotionally intense mode of communication, aligning with the findings from the LIWC analysis.

These open vocabulary results confirm Hypothesis 1: female users on Weibo exhibit mixed emotional content, whereas female Facebook users display more consistently positive emotions, especially excitement, and frequently use exclamation marks. Male Facebook users show informal and intense language, including profanity, reflecting a more direct and emotionally intense communication style compared to Weibo. The results also support *Hypothesis 2* by showing that female users on both Weibo and Facebook use more relational and emotion-related language, reflecting communal content, while male users discuss topics like sports, politics, and socio-economic issues, indicating agentic content. The findings also support *Hypothesis 3*, as female Weibo users frequently reference appearance and bodily looks, and male users engage in national identity, political, and

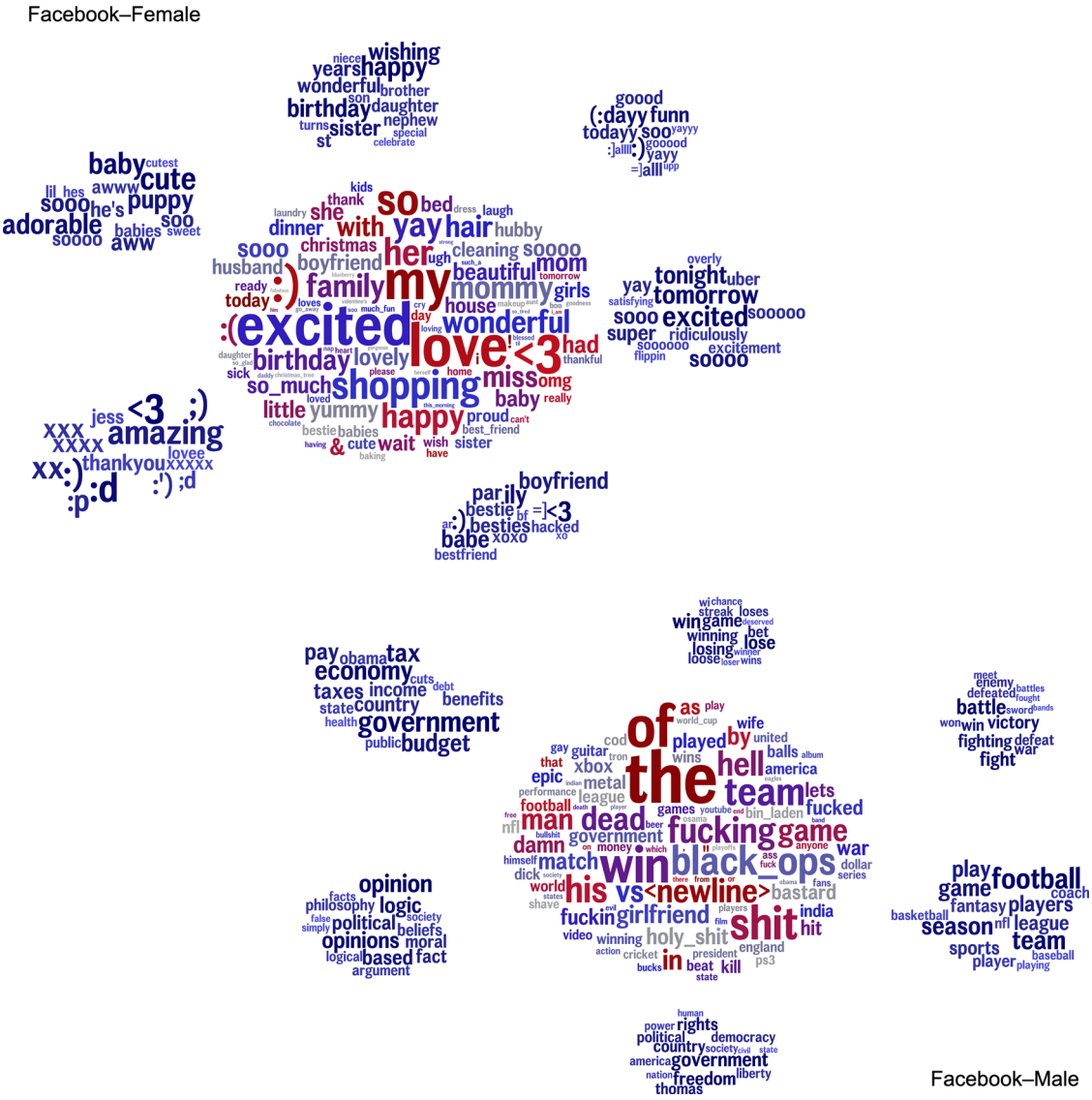

Fig. 2 Words and phrases for female and male users on Facebook. The 100 most distinctive words and phrases, as well as the latent Dirichlet allocation topics for each gender on Facebook. Results of female users are on the top left, and results of male users are on the bottom right. Correlations were adjusted for age and statistically significant ($p < 0.05$, two-tailed *t*-test, Benjamini–Hochberg corrected).

socio-economic discussions, reinforcing traditional masculine roles.

Age

Closed vocabulary. Our correlational analysis revealed distinct age-related language patterns on Weibo and Facebook (see Table 3). As age increased, Weibo users' postings were more associated with *work*, *money*, and *achievement*, while as age decreased, Weibo users were more likely to express their emotions and attitudes. Top-correlated categories for younger age were *informal language* (including *assent*, *swearing* words, and *nonfluencies*), *affective processes* (including *positive emotions* and *negative emotions* such as *anxiety*), and *first personal singular pronouns*. Older users on Weibo, on the other hand, talked more about *personal concerns* (including *work*, *money* and *home*), their *drives* (*achievement*). They also used more *function words*, such as *prepositions* and *conjunctions*, as well as quantitative words, such as *space* and *quantity units*, in their posts. Furthermore, older users used more impersonal pronouns and mentioned social processes and health relatively often. Older users more often shared what they are eating (*Ingestion*) while younger users more often shared what they are experiencing (*hear* and *feel*). Older users had more of a *present focus*, while younger users had more of a *future focus*.

Similar to the Weibo findings, the top-correlated LIWC categories associated with older age on Facebook are personal concerns such as *money*, *achievement*, *power* and social processes such as *family*, while the top-correlated LIWC categories associated with younger age are more about self in that they use *first personal singular pronouns* quite often.

To illustrate the curvilinear (U-shaped) pattern of positive emotions with age, we categorized participants into age subgroups: 13–18, 19–22, 23–29, 30–44, 45–55, and 55+ years (see Fig. 3). The 55+ group was analyzed separately, as literature indicates that the U-turn in positive emotions typically occurs in middle age. As visually indicated in Fig. 3, the use of positive emotion language decreased from adolescence through midlife and appeared to rise again in the 55+ group on both Weibo and Facebook—suggesting a descriptive U-shaped pattern. However, this observation is based on descriptive trends rather than formal

Table 3 The top 15 most positively and negatively correlated linguistic inquiry and word count (LIWC) categories with age on Weibo and Facebook.

Weibo						Facebook					
Supercategory	Subcategory	Representative words	95% CI			Supercategory	Subcategory	Representative words	95% CI		
			r	L	U				*r*	L	U
Positively correlated											
Personal concerns	Work	工厂, 面试, 薪水 (factory, interview, salary)	0.215	0.194	0.235	Function words	Articles	a, an, the	0.154	0.134	0.175
	Money	富有, 年薪, 折扣 (wealth, annual salary, discount)	0.159	0.139	0.180		Prepositions	to, with, above	0.140	0.119	0.161
	Home	房子, 家人, 宠物 (house, family, pets)	0.073	0.052	0.094	Personal concerns	Money	audit, cash, owe	0.149	0.128	0.169
Drives	Achievement	擅长, 尽责, 高手 (be good at, responsible, master)	0.169	0.148	0.189		Home	kitchen, landlord	0.128	0.108	0.149
Function words	Prepositions	从, 依照, 把 (from, follow, refer to give/take something)	0.153	0.132	0.174		Religion	altar, church	0.114	0.094	0.135
	Conjunctions	和, 一旦, 不仅 (and, in case, not only)	0.095	0.074	0.116	Personal pronouns	3rd person plural	they, their, they'd	0.144	0.123	0.164
	Multiple function[c]	的, 有, 是 (of, have, is)	0.092	0.071	0.112		3rd person singular	she, her, him	0.076	0.055	0.097
	Impersonal pronouns	一切, 这些, 其他 (everything, these, other)	0.080	0.059	0.101	Social processes	Family	daughter, dad, aunt	0.128	0.107	0.148
Cognitive processes	Causation	引起, 使得, 变成(cause, make, become)	0.145	0.124	0.165	Other Grammar	Quantifiers	few, many, much	0.124	0.103	0.145
Social processes	Humans[c]	人民, 成员, 群众(people, members, masses)	0.088	0.068	0.109	Drives	Power	superior, bully	0.123	0.103	0.144
Relativity	Space	里面, 街道, 台上 (inside, street, on stage)	0.088	0.067	0.109		Achievement	win, success, better	0.091	0.070	0.111
	Postposition[c]	之中, 以上, 为止 (among, above, till)	0.081	0.060	0.102		Affiliation	ally, friend, social	0.078	0.057	0.099
Chinese grammar	Quantity unit[c]	条, 头, 枝 (strips, heads, branches)	0.087	0.066	0.108	Relativity	Space	down, in, thin	0.107	0.086	0.128
Tense marker	Present tense marker[c]	目前, 现今, 平时(currently, now, usually)	0.083	0.062	0.104		Relativity	area, bend, exit	0.085	0.064	0.105
Biological processes	Ingestion	消化, 吃, 煮菜 (digest, eat, cook)	0.077	0.056	0.098	Biological processes	Ingestion	dish, eat, pizza	0.091	0.071	0.112
Negatively correlated											
Informal language	Assent	了解, 真的, 好(understand, really, good)	−0.235	−0.255	−0.215	Informal language	Netspeak	btw, lol, thx	−0.188	−0.208	−0.168
	Swear words	去你的, 智障, 机车[b] (fuck you, mentally retarded, annoying)	−0.086	−0.107	−0.065		Fillers	Imean, youknow	−0.063	−0.084	−0.042

Table 3 (continued)

Weibo						Facebook					
Supercategory	**Subcategory**	**Representative words**	**95% CI**			**Supercategory**	**Subcategory**	**Representative words**	**95% CI**		
			r	**L**	**U**				***r***	**L**	**U**
	Nonfluencies	呃, 然后, 那 (uh, then, that)	−0.159	−0.179	−0.138		Swear words	fuck, damn, shit	−0.119	−0.140	−0.098
Affective processes	Positive emotion	信心, 满足, 祝福 (confidence, satisfaction, blessing)	−0.195	−0.215	−0.174		Assent	agree, OK, yes	−0.097	−0.118	−0.076
	Anxiety	不安, 挣扎, 紧绷 (restless, struggling, tight)	−0.073	−0.094	−0.052		Nonfluencies	er, hm, umm	−0.084	−0.105	−0.063
Personal pronouns	1st person singular	本人, 自己, 我 (me, myself, I)	−0.194	−0.214	−0.174	Affective processes	Negative emotion	hurt, ugly, nasty	−0.183	−0.203	−0.162
	2nd person singular	你, 妳, 您 (you)	−0.125	−0.146	−0.105		Anger	hate, kill, annoyed	−0.144	−0.164	−0.123
Tense marker[c]	Progress tense marker[c]	了, 至今, 近期 (have done, so far, recently)	−0.116	−0.137	−0.095		Sadness	crying, grief, sad	−0.085	−0.106	−0.064
	Future tense marker[c]	之后, 即将, 将来 (after, about, in the future)	−0.102	−0.123	−0.081	Personal pronouns	1st person singular	I, me, mine	−0.179	−0.199	−0.159
Perceptual processes	Hear	大叫, 听见, 谈话 (shout, hear, talk)	−0.100	−0.121	−0.079	Biological processes	Sexual	horny, love, incest	−0.096	−0.117	−0.076
	Feel	平滑, 弹性, 触摸 (smooth, elastic, touch)	−0.038	−0.059	−0.017		Body	cheek, hands, spit	−0.075	−0.096	−0.054
Cognitive process	Insight	了解, 恍然大悟, 体会 (understand, realize, experience)	−0.099	−0.120	−0.078	Function words	Negations	no, not, never	−0.095	−0.115	−0.074
Chinese grammar	Interjunction[c]	呢, 吗, 吧[a] (*ne, ma, ba*)	−0.086	−0.107	−0.065		Common Adverbs	very, really	−0.090	−0.111	−0.069
Biological processes	Body	脖子, 皮肤, 肠胃 (neck, skin, stomach)	−0.076	−0.097	−0.055	Personal concerns	Death	bury, coffin, kill	−0.091	−0.112	−0.070
Social processes	Friend	同伴, 朋友, 麻吉[b] (companion, friend, best friend)	−0.032	−0.053	−0.011	Perceptual processes	Hear	listen, hearing	−0.072	−0.093	−0.051

LIWC category correlations with age across the two platforms. 95% CI = 95% confidence interval, L = Lower, U = Upper; *r* = standardized linear regression coefficients.
All correlations in the table were significant ($p < 0.005$; BH *p*-corrected). For interpretability, subcategories are grouped under their LIWC supercategories based on the English LIWC structure.
[a]Markers for the perfective aspect in standard Chinese grammar expressing completion.
[b]A more typical phrase used in Taiwan than in Mainland China.
[c]Category unique to Chinese LIWC, with no English equivalent.

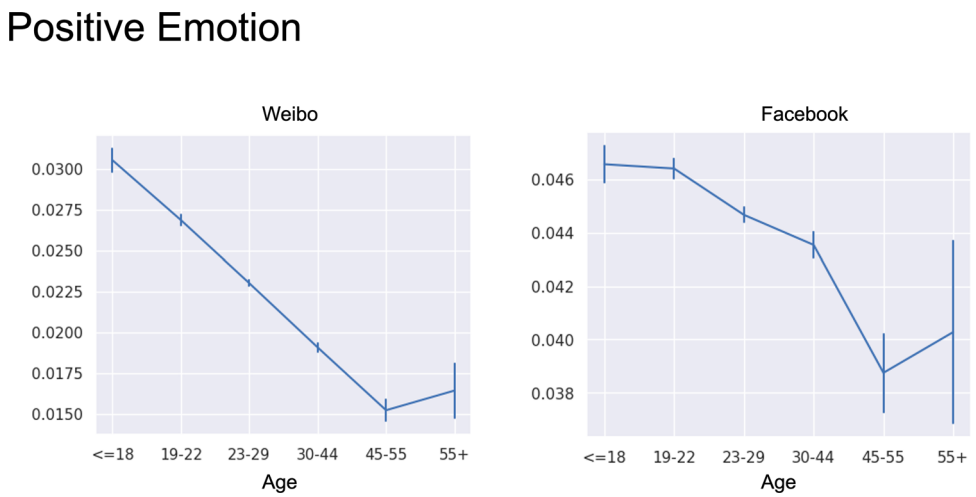

Fig. 3 Use of positive emotion words across the lifespan by age groups. *Y*-axis is the normalized frequency of words per user. Each bar represents the mean value of the respective LIWC variable for each age group, with error bars indicating standard error (SE). *X*-axis is the age groups of the users: 13–18 ($n = 764$), 19–22 ($n = 2475$), 23–29 ($n = 4176$), 30–44 ($n = 1194$), 45–55 ($n = 90$), and 55+ years ($n = 29$).

statistical testing (e.g., polynomial regression) and should not be interpreted as conclusive evidence of a curvilinear relationship. We caution against overinterpreting this uptick in the oldest group, as the 55+ category contained relatively few participants ($n = 29$ per platform), and the observed increase may be influenced by sample size and variability.

In line with *Hypothesis 4*, our closed-vocabulary suggests a general trend in which positive emotion words decrease until midlife and appear to increase again in later adulthood. The findings also support *Hypothesis 5:* with age, individuals on both platforms used fewer first-person singular pronouns ('I') and more social words, which suggests a shift away from a self-focused perspective toward socially meaningful goals.

Open vocabulary. The 100 most distinctive words and phrases as well as the LDA topics for each age group on Weibo (Fig. 4) and Facebook (Fig. 5) are visualized (see Supplementary Fig. S2 for a visualization in the original Chinese language). In general, we could observe a trend shift from exam-related topics to personal life and work, and then to politics and history topics.

The 13–18-year-old Weibo users posted more often about emotions (specifically positive and surprised sighs, e.g., *haha-haha*, *Ahhhhh*), school (*senior year of high school*, *college-entrance exam*), informal language (*Haha*), celebrities (*Han Lu*), polite version of slang (*wo cao—translated as wtf*). They also said lots of *goodbyes* (*bye-bye*), which is a direct translation of bye from English instead of using zai jian, suggesting the adoption of Westernized expressions. For Facebook users in the 13 to 18 age group, the word cloud also showed an emphasis on school life (with frequent mentions of *school* and *homework*) and positive emotions (*amazing*, *fun*). What stood out differently about Weibo is the frequent mentions of social relationships (specifically romantic relationships such as *boyfriend*, *love*, *crush*, *single*).

The 19–22-year-old Weibo users also often said *goodbyes* and expressed emotions and feelings, both positive and negative (*love you*, *awesome*, *frustrated*). Unlike their younger counterparts, they talked more about college/university life (*campus*, *college*) instead of school life and exams. Furthermore, they seem to care a lot about the physical appearance (*handsome*, *ugly*, *face*). On Facebook, the 19 to 22 group also talked about college life, but continued the trend of social focus (*single*, *kiss*) and positive emotions (*funn*, *yayyy*).

Words and phrases most associated with 23–29-year-old users on both platforms are work such as *job* and *work* (more so on Facebook). While on Weibo, users talked about quite a bit about relationship (*marriage*, *men* and *women*) and food, Facebook users were more interested in drinking (*drink* and *beer*) and expressed more positive emotions (*happy* and *excited*).

The 30–44-year-old Weibo users shared mostly topics of national and political significance, including *politics*, *society* and *policies*, *market* and *economy*, and *safety*. Additionally, words such as *China* signal a strong engagement with national identity. The mention of *children* suggested a balance between family and societal responsibilities. In contrast, users on Facebook in this age group talked almost exclusively about family and social ties, with words such as *kids*, *family*, *lunch* and friends, suggesting a focus on family and social life.

Lastly, the 45–64-year-old Weibo users posted more about *history*, *politics*, *government* and *legal agencies*, *news platforms* and *media*, and *policies*, implying an engagement with social issues and media. Comparatively, users of the same age group on Facebook also talked about *politics*, *government*, and *society*. Beyond these topics, they also mentioned social ties (*family* and *friends*), spirituality (*praying*), and health (*cancer* and *patient*), potentially reflecting broader life concerns and values.

The open vocabulary findings support *hypothesis 5* by showing a shift from self-focused topics (e.g., personal achievements) to socially meaningful themes (e.g., family and societal issues) as users age on both platforms. The findings further support *hypothesis 5* by showing that older Weibo users engage with collectivistic themes, such as national identity and societal issues, reflecting collectivist tendencies. In contrast, older Facebook users focus on personal family and social life.

Discussion

This study highlights key findings that advance our understanding of how gender and age-related language patterns are shaped by cultural contexts, as revealed through naturalistic social media data. Using social role theory and socioemotional selectivity theory as guiding frameworks, we explored universal and culturally specific linguistic trends across gender and age on Weibo and Facebook. Our findings confirm and extend these theories, offering new insights into how societal norms and psychological development manifest in digital communication.

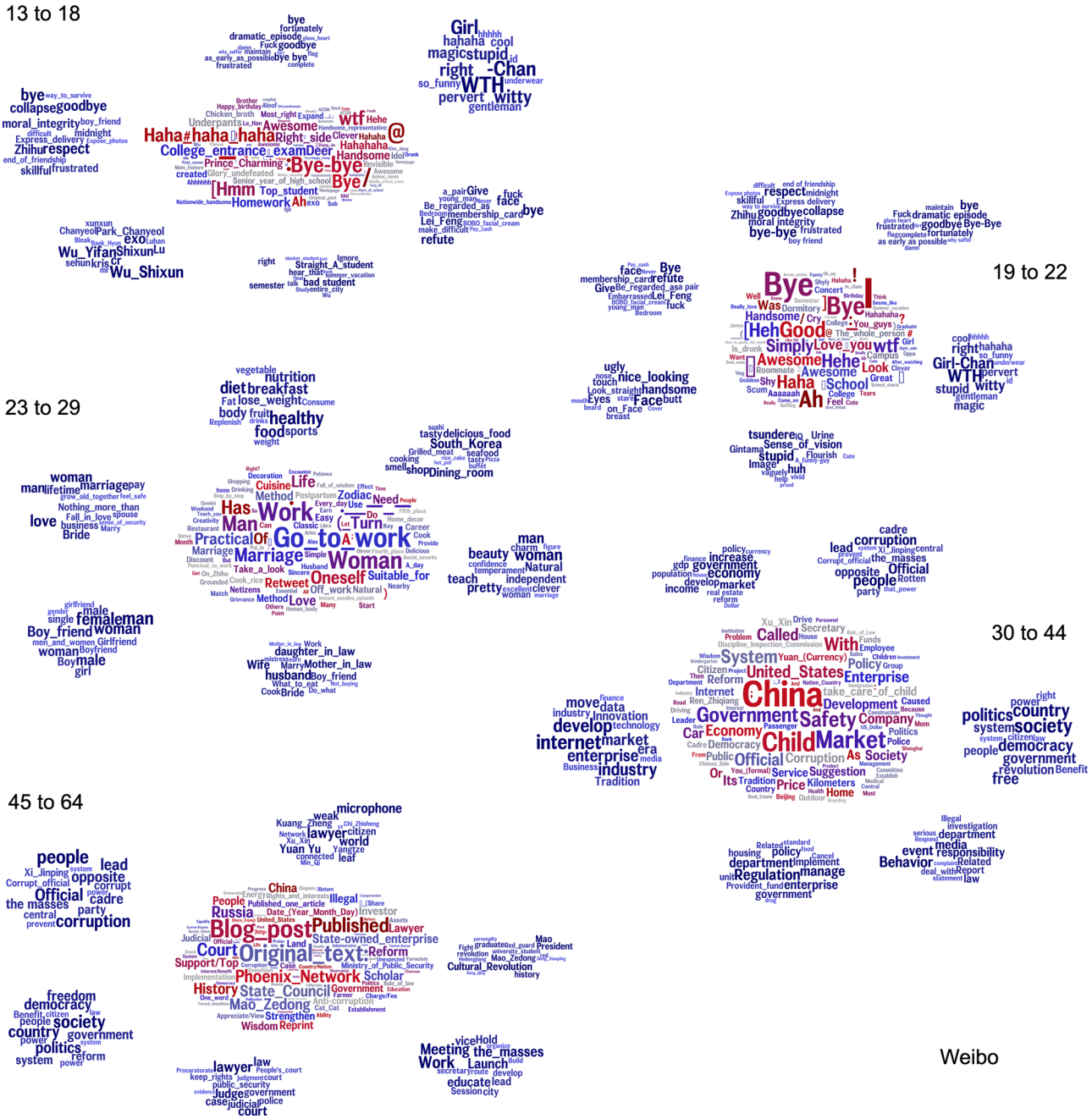

Fig. 4 Words and topics distinguished Weibo users by age. Words and topics that most highly distinguished Weibo users aged 13 to 18, 19 to 22, 23 to 29, 30 to 44, 45 to 64. Ordered from top to bottom: 13 to 18, 19 to 22, 23 to 29, 30 to 44, 45 to 64. Correlations were adjusted for gender and statistically significant ($p < 0.05$, two-tailed t-test, Benjamini–Hochberg corrected).

Theoretical implications

Gender differences and cultural norms. Female Weibo users demonstrated a balance of positive and negative emotions, reflecting the East Asian cultural tendency to value emotional moderation and the integration of both positive and negative states, often seen as a sign of emotional maturity (Josefsson et al., 2013). This mixed emotional expression contrasts with female Facebook users, who displayed higher emotional arousal with predominantly positive expressions and celebratory language, such as excitement and enthusiasm, supported by frequent exclamation marks. This language style aligns with Western norms that prioritize maximizing positive states and emphasize high-arousal emotions (Sims et al., 2015; Tsai et al., 2006), revealing a cultural preference for openly expressing positivity. Similarly, male Weibo users tended to engage more in national and political discourse, suggesting that male users on Weibo feel a greater societal role to express engagement with collective matters, aligning with East Asian collectivistic values that emphasize group identity and social responsibility. In contrast, male Facebook users displayed more informal and emotionally intense expressions, often using direct language, including profanity, suggesting a culturally permissible form of emotional directness and informality characteristic of Western communication styles (Wang, 2022).

Our findings provide support for social role theory by illustrating how gender differences in language use reflect societal roles and cultural norms. As SRT posits, gendered behavior and language patterns are deeply rooted in the division of labor and

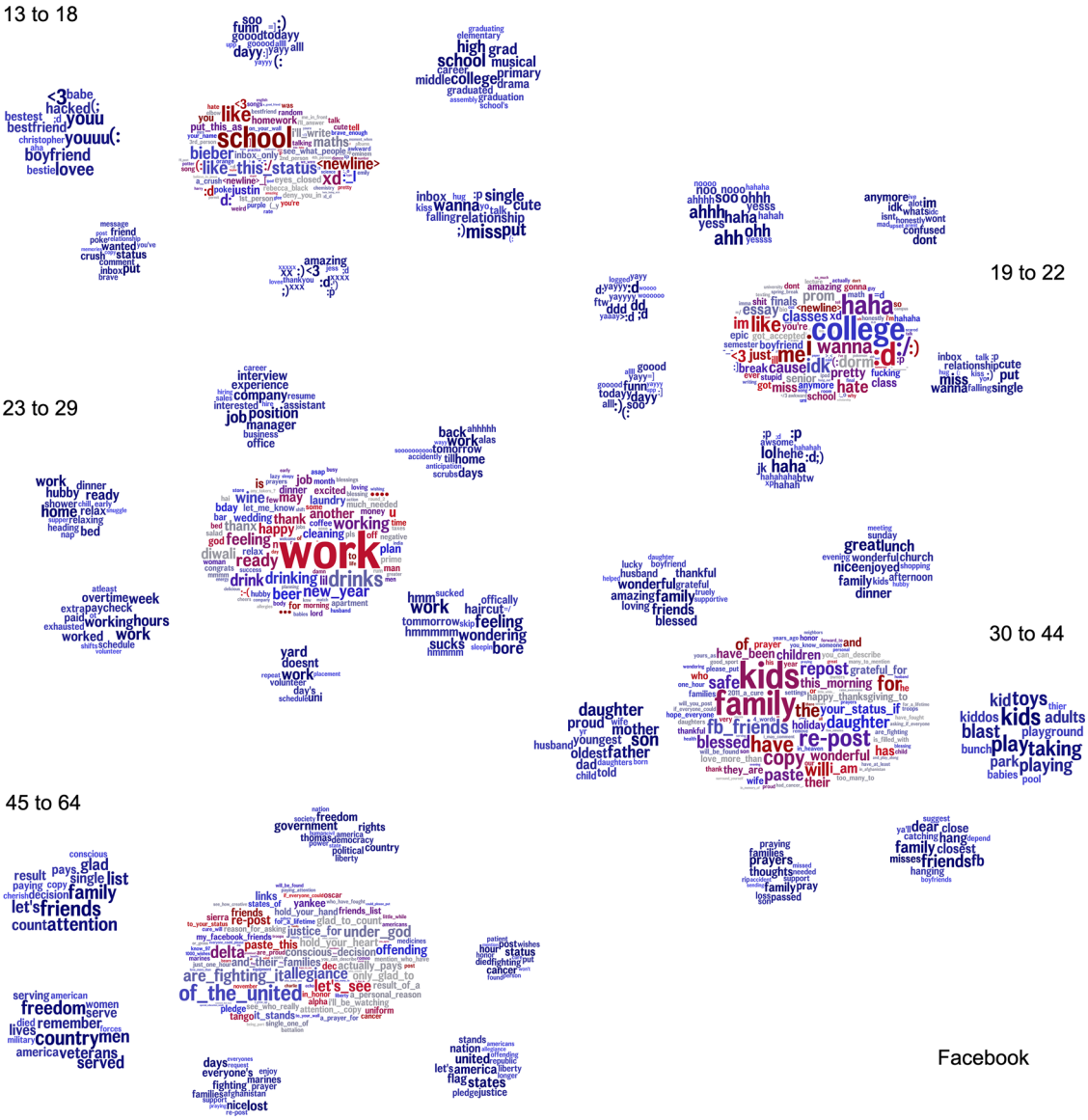

Fig. 5 Words and topics distinguished Facebook users by age. Words and topics that most highly distinguished Facebook users aged 13 to 18, 19 to 22, 23 to 29, 30 to 44, 45 to 64. Ordered from top to bottom: 13 to 18, 19 to 22, 23 to 29, 30 to 44, 45 to 64. Correlations were adjusted for gender and statistically significant ($p < 0.05$, two-tailed *t*-test, Benjamini–Hochberg corrected).

reinforced by societal expectations and stereotypes (Martin and Ruble, 2010; Wood and Eagly, 2015). Across both platforms, women used more communal and relational language, while men focused on agentic and task-oriented topics, aligning with traditional roles that associate women with nurturing and relational responsibilities and men with independence and achievement.

Importantly, our findings also extend SRT by demonstrating how cultural contexts shape the expression of these gender roles. On Weibo, a platform rooted in East Asian collectivism, gender roles appeared more pronounced, with women's language reflecting traits associated with "kawaii" (cuteness) ideals, such as terms like 'cute' and 'small'. These traits align with Confucian cultural values that emphasize modesty and emotional restraint in women, reinforcing traditional femininity. In contrast, men on Weibo engaged heavily in discussions about politics and national identity, reflecting collectivist values that associate masculinity with societal responsibility and leadership (Gao, 2012; Rosenlee, 2012). However, we recognize that the observed "distinctive profiles" may both reflect and reinforce cultural expectations. Whether these language patterns are primarily shaped by socialization and cultural context within the unique environment of social media, or if they reflect inherent gender differences, is beyond the scope of this study and remains unanswered.

Age-related language shifts embedded in cultural contexts. Our findings provide support for socioemotional selectivity theory by illustrating how age-related shifts are reflected in language use. Specifically, we observed a U-shaped trend in the use of positive emotion words across age groups, with a decline during midlife and a potential increase in later adulthood. This pattern is consistent with SST, which posits that as individuals age, they prioritize emotionally meaningful goals (Carstensen et al., 2003). Similar findings in other studies confirm this pattern across different cultures, emphasizing a midlife dip in well-being followed by a later-life increase (Blanchflower and Graham, 2022). However, the observed increase in positive emotion among the oldest group (55+) should be interpreted cautiously, given the relatively small sample size and greater variability in this subgroup. Additionally, the decreasing use of first-person singular pronouns ('I')

with age and the simultaneous increase in social words align with SST's premise that older individuals focus on socially meaningful goals.

Distinct thematic engagement across life stages further supports and extends SST's framework by embedding it in cultural contexts. For younger users (ages 13–22) on both platforms, discussions centered around school, exams, college life, and emotional expressions, aligning with this stage's focus on academic achievement, social interactions, and identity formation. As users transitioned into early adulthood (ages 23–29), topics shifted toward work and relationships, reflecting goal-driven priorities characteristic of this life stage.

For middle-aged users (ages 30–44), the focus diverged: Weibo users shifted toward political and national discourse, consistent with collectivist values emphasizing broader societal roles, while Facebook users centered on family and personal achievements, reflecting individualist values. Marriage and children emerged as topics earlier on Weibo than on Facebook, likely reflecting sociocultural differences, including a lower median age of marriage in China compared to the United States at the time of data collection. On Weibo, family discussions often highlighted dynamics within extended families, particularly the daughter-in-law and mother-in-law relationship ('婆媳关系'), reflecting a cultural emphasis on maintaining harmony within a collective family structure. Conversely, Facebook users focused more on individual relationships within the nuclear family, frequently mentioning specific family members like mothers and daughters, consistent with Western norms of family individualism.

Older Weibo users (45+) are particularly engaged in discussions centered around national identity, political matters, and social issues, reflecting the collectivist and politically involved nature prevalent in East Asian cultures. Conversely, older Facebook users focused more on family, health, and spirituality, underscoring a Western emphasis on personal well-being, individual fulfillment, and close social ties in later life stages. This contrast may have deep roots in Confucian philosophy, which profoundly shapes Chinese cultural values (Rozi, 2020).

Confucian teachings (Rozi, 2020) emphasize a progression of personal and societal responsibilities across life stages, encapsulated in ideals such as '三十而立' (one should become independent at thirty) and '修身, 齐家, 治国, 平天下' (cultivate oneself, regulate one's family, govern the nation, and bring peace to the world). These teachings stress that achieving personal development and assuming social responsibility, especially for men, is paramount in adulthood. Engaging in discourse on national and societal issues provides a practical means for many older Weibo users to embody these values, addressing societal concerns as an extension of self-cultivation and duty. This pattern contrasts with Facebook's Western-leaning cultural environment, where a focus on individual relationships, personal health, and spiritual contentment typically becomes more prominent with age, reflecting a life course oriented toward personal and familial well-being rather than broader social engagement.

Cultural convergence. Our analysis revealed that younger users across both platforms tend to show more linguistic similarities than older users, likely reflecting generational shifts influenced by globalization and digital interconnectedness. Particularly in China, younger generations born after the 1980s have experienced rapid economic growth, modernization, and increased exposure to global media and cultural trends. This has allowed for shared experiences in language and cultural references across borders. For instance, a teenage girl in China may follow the same Korean boy bands or popular television series as her American counterpart. This cultural convergence illustrates the merging of Western and Eastern influences, shaping not only language use but also a broader array of cultural expressions. This raises the question: Is cultural norm (e.g., Confucianism) adapting to modern influences, or is it losing ground to individualistic ideals? The potential interplay between Westernization and Confucianism represents a critical area for future research, particularly as younger generations navigate the tension between global modernity and cultural heritage.

Implications for theory-building. This study contributes to theory-building in two important ways. First, it extends social role theory by demonstrating that gendered language is not only shaped by societal expectations but also amplified or attenuated through culturally specific online environments. The more pronounced gender differences on Weibo suggest that collectivist and tight cultural norms may reinforce traditional roles more strongly than individualist, loose cultures—adding a cross-cultural layer to SRT. Second, our findings support and refine socioemotional selectivity theory by illustrating how language priorities shift with age in culturally nuanced ways. For instance, older users on Weibo engaged more with national identity and social responsibility, while their Facebook counterparts emphasized emotional well-being and spirituality. This suggests that while the age-related motivational shift toward meaningful goals is universal, the content of these goals is culturally framed—offering a culturally expanded understanding of SST.

Practical implications. These findings carry important practical implications. First, they offer insight into designing culturally sensitive AI models for sentiment analysis, psychological assessments, or mental health coaches (Rai et al., 2024). Language indicators of gender and age differ across cultures—not only in content but also in emotional tone and intensity—suggesting that cross-cultural adaptation is essential for fairness and accuracy in digital tools. For instance, a positivity classifier trained on Western data may misinterpret East Asian mixed-emotion language as neutral or even negative, underestimating emotional expression.

Second, our findings can inform cross-cultural communication strategies in multinational organizations, education, or diplomacy. Understanding how cultural values shape emotional expression and social focus can improve empathy, reduce miscommunication, and enhance message tailoring in global teams. For example, leadership communication may need to be adjusted to account for emotional restraint in East Asian contexts versus emotional directness in Western ones (Rai et al., 2025).

Finally, the generational convergence we observed—particularly among younger users—suggests the potential for shared cultural touchpoints that transcend national boundaries. However, this convergence also raises questions about the preservation of cultural heritage, particularly in rapidly modernizing societies. Policymakers, educators, and platform designers alike should consider how digital spaces can support both global connection and cultural continuity.

Limitations. While the findings of this study demonstrate strong face validity in the language results and fair predictive accuracy, several limitations must be acknowledged. First, while our findings reflect established gendered language patterns, we acknowledge that these may be influenced by the cultural and social contexts specific to our sample and platforms. Consequently, our study may capture social expectations about gender rather than purely intrinsic traits. Furthermore, we recognize that social media itself may amplify stereotypical gender norms, as users may subconsciously or strategically adopt language patterns

aligned with societal expectations for their gender. This dynamic could limit the extent to which our findings represent broader, context-independent traits, as opposed to platform-specific behaviors shaped by audience and social pressures. As such, these findings should be interpreted as reflective of language use within specific social media contexts rather than universal gender characteristics.

Second, our analysis may be impacted by selection bias inherent to the user demographics of Weibo and Facebook. Research indicates that users of both platforms tend to be younger, wealthier, and better educated than the national averages (Greenwood et al., 2016; Wang, 2018). This age skew is also reflected in our dataset (see Table 1). In China, approximately 50% of netizens utilize Weibo (Wang, 2018), which, similar to Facebook's user base in the United States, may not represent the broader population effectively. Furthermore, although the Weibo sample was drawn using a randomized selection approach, this does not guarantee representativeness. The Facebook sample, in contrast, stems from voluntary participation in the myPersonality project, adding another layer of self-selection bias. As such, both samples may overrepresent digitally engaged, socially expressive users within their respective cultures. This demographic skew raises concerns about the generalizability of our findings, as the unique characteristics of this sample may influence the language and topics discussed. For instance, Facebook users are often more politically liberal, which may affect their linguistic choices (Greenwood et al., 2016). Furthermore, the older generation (ages 45+) in China has experienced significant social and economic changes, leading to variability in education, income, and lifestyle. Therefore, the differences observed in language use across age groups on these platforms may not solely reflect cultural disparities but could also be a result of selection bias. As such, our findings should be interpreted as reflective of active, digitally engaged social media cohorts in China and the United States, rather than as generalizable insights about their entire national populations. We also acknowledge a temporal mismatch between datasets—Facebook data were collected between 2010 and 2011, and Weibo data in 2014. Although this gap may introduce minor contextual variation, we found no evidence of systematic linguistic shifts attributable to time. Since our focus lies on broad demographic and cultural patterns, we believe the impact of this temporal difference is limited.

Third, the public dataset from Weibo may not fully capture the effects of gender and age due to internet censorship in China (Vuori and Paltemaa, 2015). For instance, language associated with swearing or sensitive political topics may be less prevalent on Weibo due to Cyber Ecology Governance Regulations in China (Central Cyberspace Affairs Commission, 2019; Paltemaa et al., 2020).

Fourth, although social media platforms provide access to vast user bases, the cohorts analyzed in this study are relatively small compared to the total population of social media users. This segmentation leads to further subdivision for linguistic analysis, resulting in limited statistical power. Larger cohort sizes—ideally 10 to 100 times greater—could provide more robust findings and mitigate biases related to the comparatively small sample sizes in our study, particularly for the 45+ age group.

Finally, the cross-sectional nature of our data further restricts our ability to track individual users' language evolution over time, which would enhance our understanding of how language changes with age.

Conclusion

This study affirmed and expanded social role theory and socioemotional selectivity theory by analyzing gender- and age-related language on culturally distinct social media platforms, Weibo and Facebook. Using SRT, we confirmed universal gender patterns—women used more communal language, and men focused on agentic themes—while highlighting cultural modulation. Weibo users adhered more strictly to traditional gender roles, reflecting collectivist and individualist norms, respectively. For SST, we validated its prediction that aging shifts priorities toward emotionally meaningful and socially significant goals. While the overall trajectory of positive emotional expression appeared U-shaped, cultural differences emerged in thematic focus: older Weibo users focused on societal themes like national identity, while Facebook users emphasized personal well-being and family. These findings demonstrate how culture shapes universal psychological processes, advancing both theories by integrating cross-cultural perspectives. By leveraging social media as a naturalistic dataset, we provided ecologically valid insights into how societal norms influence emotional and social priorities across life stages.

Data availability

Given the proprietary and sensitive nature of the raw Facebook profiles, we do not publicly release these data and only make them available upon reasonable request from the corresponding author.

Code availability

Code used in the study is publicly available at dlatk.wwbp.org.

Received: 25 January 2022; Accepted: 4 September 2025;
Published online: 31 October 2025

Notes

1 Some Weibo users reported unrealistic ages, such as 120+, and therefore were excluded from the dataset.

2 The term "strongly" or "weakly" used in the present paper to indicate the degree of the associations is relative and does not refer to the effect size of the correspondent correlations.

References

Blanchflower DG, Graham CL (2022) The mid-life dip in well-being: a critique. Soc Indic Res 161:287–344

Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. J Mach Learn Res 3:993–1022

Burger M, Thornborrow J, Fitzgerald R (2017) Analyzing new media and social media as interactive spaces. In M Burger (ed),Discours des réseaux sociaux: Enjeux publics, politiques et médiatiques. De Boeck Supérieur, pp. 25-40. https://doi.org/10.3917/dbu.thorn.2017.01.0025

Carstensen LL, Fung HH, Charles ST (2003) Socioemotional selectivity theory and the regulation of emotion in the second half of life. Motiv Emot 27:103–123. https://doi.org/10.1023/A:1024569803230

Carstensen LL, Isaacowitz DM, Charles ST (1999) Taking time seriously: a theory of socioemotional selectivity. Am Psychol 54:165–181

Central Cyberspace Affairs Commission (2019) 网络生态治理规定(征求意见稿) [Cyber ecology governance regulations (draft for comment)]. Available via https://tech.sina.com.cn/i/2019-09-10/doc-iicezzrq4866232.shtml

Chiu CY, Leung AK-Y, Hong YY (2010) Cultural processes: an overview. In: Leung AK-Y, Chiu CY, Hong YY (eds) Cultural processes: a social psychological perspective. Cambridge University Press, Cambridge, pp 3–22

Cho YM, Pang D, Thapa S, Sherman G, Ungar LH, Tay L, Guntuku SC (2025) Language-based valence and arousal expressions between the United States and China: a cross-cultural examination. In: Chiruzzo L, Ritter A, Wang L (eds) Findings of the Association for Computational Linguistics: NAACL 2025: Albuquerque, April 2025, pp 5586–5600

Church K, Hanks P (2002) Word association norms, mutual information, and lexicography. Comput Linguist 16:22–29

Cui J, Zhang T, Jaidka K, Pang D, Sherman G, Jakhetiya V, Ungar LH, Guntuku SC (2022) Social media reveals urban-rural differences in stress across China. In: Proceedings of the international AAAI conference on Web and social media, Atlanta, June 2022, pp 114–124

Cohen J (1988) Set correlation and contingency tables. Appl Psychol Meas 12:425-434. https://doi.org/10.1177/014662168801200410
Dixon EL, Joshi SM, Ferrell W, Volpp KG, Merchant RM, Guntuku SC (2022) COVID-19 contact tracing app reviews reveal concerns and motivations around adoption. PLoS ONE 17(9):e0273222
Eagly AH, Nater C, Miller DI et al. (2020) Gender stereotypes have changed: a cross-temporal meta-analysis of U.S. public opinion polls from 1946 to 2018. Am Psychol 75:301–315
Fishman JA (1980) Bilingualism and biculturism as individual and as societal phenomena. J Multiling Multicult 1:3–15
Fu G, Xu F, Cameron CA et al. (2007) Cross-cultural differences in children's choices, categorizations, and evaluations of truths and lies. Dev Psychol 43:278–293
Gao E, Zuo X, Wang L et al. (2012) How does traditional Confucian culture influence adolescents' sexual behavior in three Asian cities? J Adolesc Health 50:S12–S17
Garimella A, Mihalcea R, Pennebaker J (2016) Identifying cross-cultural differences in word usage. In: Matsumoto Y, R Prasad (eds) Proceedings of COLING 2016, the 26th international conference on computational linguistics: technical papers, Osaka, December 2016. The COLING 2016 Organizing Committee, pp 674–683
Gelfand MJ, Raver JL, Nishii L et al. (2011) Differences between tight and loose cultures: a 33-nation study. Science 332:1100–1104
Greenwood S, Perrin A, Duggan M (2016) Social media update 2016. Pew Research Center. Available via http://assets.pewresearch.org/wp-content/uploads/sites/14/2016/11/10132827/PI_2016.11.11_Social-Media-Update_FINAL.pdf. Accessed 4 Nov 2024
Guntuku SC, Li M, Tay L, Ungar LH (2019) Studying cultural differences in emoji usage across the east and the west. In: Proceedings of the international AAAI conference on Web and social media. Munich, June 2019, pp 226–235
Hofstede G (1980) Motivation, leadership, and organization: do American theories apply abroad? Organ Dyn 9:42–63
Hofstede G (2001) Culture's recent consequences: using dimension scores in theory and research. Int J Cross Cult Manag 1:11–17
Hsu N, Badura KL, Newman DA, Speach MEP (2021) Gender, "masculinity," and "femininity": a meta-analytic review of gender differences in agency and communion. Psychol Bull 147:987–1011
Huang CL, Chung CK, Hui N, Lin YC, Seih YT, Chen WC, Lam B, Bond M, Pennebaker JW (2012) The development of the Chinese linguistic inquiry and word count dictionary. Chin J Psychol 54(2):185–201
Jaidka K, Chhaya N, Ungar LH (2018) Diachronic degradation of language models: Insights from social media. In Proceedings of the 56th Annual Meeting of the Association for Computational, Melbourne, July 2018, pp 195–200. https://doi.org/10.18653/v1/P18-2032
Josefsson K, Jokela M, Cloninger CR et al. (2013) Maturity and change in personality: developmental trends of temperament and character in adulthood. Dev Psychopathol 25:713–727
Joshi PD, Wakslak CJ, Appel G, Huang L (2020) Gender differences in communicative abstraction. J Pers Soc Psychol 118:417–435
Ju B, Sandel TL, Fitzgerald R (2019) Understanding Chinese internet and social media: the innovative and creative affordances of technology, language and culture. Cah Cent Ling Sci Lang 59:161–177
Kern ML, Eichstaedt JC, Schwartz HA et al. (2014) The online social self: an open vocabulary approach to personality. Assessment 21:158–169
Kim U, Triandis HC, Kâğitçibaşi Ç et al. (1994) Individualism and collectivism: theory, method, and applications. Sage, Thousand Oaks
Labouvie-Vief G, Diehl M, Tarnowski A, Shen J (2000) Age differences in adult personality: findings from the United States and China. J Gerontol B Psychol Sci Soc Sci 55(1):4–17
Lim N (2016) Cultural differences in emotion: differences in emotional arousal level between the East and the West. Integr Med Res 5(2):105–109
Lui M, Baldwin T, Zhang M (2012) langid.py: an off-the-shelf language identification tool. In: Proceedings of the 50th annual meeting of the Association for Computational Linguistics, Jeju Island, 8–14 July 2012
Liu WM, Iwamoto DK (2006) Asian American men's gender role conflict: the role of Asian values, self-esteem, and psychological distress. Psychol Men Masc 7(3):153–164
Louie K (2002) Theorizing Chinese masculinity: society and gender in China. Cambridge University Press, Cambridge
Loveys K, Torrez J, Fine A et al. (2018) Cross-cultural differences in language markers of depression online. In: Loveys K, Niederhoffer K, Prud'hommeaux E, Resnik R, Resnik P (eds) Proceedings of the fifth workshop on computational linguistics and clinical psychology: from keyboard to clinic, New Orleans, June 2018. Association for Computational Linguistics, pp 78–87
Markus HR, Kitayama S (1991) Culture and the self: implications for cognition. Emot Motiv Psychol Rev 98:224–253
Martin CL, Ruble DN (2010) Patterns of gender development. Annu Rev Psychol 61:353–381
McAdams DP, de St. Aubin E, Logan RL (1993) Generativity among young, midlife, and older adults. Psychol Aging 8:221–230
Newman ML, Groom CJ, Handelman LD, Pennebaker JW (2008) Gender differences in language use: an analysis of 14,000 text samples. Discourse Process 45(3):211–236
Oyserman D, Lee SW (2008) Does culture influence what and how we think? Effects of priming individualism and collectivism. Psychol Bull 134(2):311–342
Paltemaa L, Vuori JA, Mattlin M, Katajisto J (2020) Meta-information censorship and the creation of the Chinanet Bubble. Inf Commun Soc 23(14):2064–2080
Papacharissi Z (2011) A networked self: identity, community, and culture on social network sites. Routledge, New York
Pennebaker JW, Booth RJ, Boyd RL, Francis ME (2015) The development and psychometric properties of LIWC2015. University of Texas at Austin, Austin
Potts C (2017) DLATK/happierfuntokenizing [GitHub]. Available via https://github.com/dlatk/happierfuntokenizing. Accessed 15 Jan 2020
Rosenlee L-HL (2012) Confucianism and women: a philosophical interpretation. State University of New York Press, Albany
Rai S, Shelat K, Jain D, Kishen A, Cho YM, Redkar M, Hardikar-Sawant S, Ungar LH, Guntuku SC (2025). Cross-cultural differences in mental health expressions on social media. In: Proceedings of the 3rd workshop on cross-cultural considerations in NLP, Albuquerque, April 2025, pp 132–142
Rai S, Stade EC, Giorgi S, Francisco A, Ungar LH, Curtis B, Guntuku SC (2024) Key language markers of depression on social media depend on race. In Proc Natl Acad Sci USA 121:e2319837121
Russell JA (1980) A circumplex model of affect. J Pers Soc Psychol 39:1161–1178. https://doi.org/10.1037/h0077714
Rozi F (2020) Confucian concept of self-cultivation and social harmony. Int J Lang Ling 7(2):129–136
Sandel TL, Buttny R, Varghese M (2019) Online interaction across three contexts: an analysis of culture and technological affordances. J Int Cult Commun Res 48(1):52–71
Schwartz HA, Eichstaedt J, Blanco E et al. (2013a) Choosing the right words: characterizing and reducing error of the word count approach. In: Proceedings of the second joint conference on lexical and computational semantics. Atlanta, June 2013
Schwartz HA, Eichstaedt J, Dziurzynski L et al. (2013b) Toward personality insights from language exploration in social media. In: AAAI spring symposium: analyzing microtext, Stanford, 15 March 2013
Schwartz HA, Eichstaedt J, Kern M et al. (2013c) Correlation values of LIWC categories with gender, age, and the five factor model of personality. PLoS ONE 8(9):e7371
Sims T, Tsai JL, Jiang D et al. (2015) Wanting to maximize the positive and minimize the negative: implications for mixed affective experience in American and Chinese contexts. J Pers Soc Psychol 109:292–315
Stillwell DJ, Kosinski M (2015) myPersonality Project. https://www.mypersonality.org. Accessed 4 Nov 2024
Sun M, Chen X, Zhang K, Guo Z, Liu Z (2016) Thunlp/THULAC (GitHub). Available via https://github.com/thunlp/THULAC. Accessed 04 Nov 2024
Tsai JL, Knutson B, Fung HH (2006) Cultural variation in affect valuation. J Pers Soc Psychol 90:288–307
Uz I (2015) The index of cultural tightness and looseness among 68 countries. J Cross Cult Psychol 46(3):319–335
Villar F (2012) Successful ageing and development: the contribution of generativity in older age. Ageing Soc 32:1087–1105
Vuori JA, Paltemaa L (2015) The lexicon of fear: Chinese internet control practice in Sina Weibo microblog censorship. Surveill Assym Ambig 13(3/4):400–421
Wang J (2018) Analysis of young Chinese users of Sina Weibo based on uses and gratifications theory. Dissertation, Jönköping University
Wang Q, Leichtman MD (2000) Same beginnings, different stories: a comparison of American and Chinese children's narratives. Child Dev 71:1329–1346
Wang Y (2022) A comparative study of communication styles between China and America. Acad J Hum Soc Sci 5:91–101
Wood W, Eagly AH (2002) A cross-cultural analysis of the behavior of women and men: implications for the origins of sex differences. Psychol Bull 128:699–727
Wood W, Eagly AH (2015) Two traditions of research on gender identity. Sex Roles 73:461–473
World Economic Forum (2020) Global gender gap report 2020. Available via http://www3.weforum.org/docs/WEF_GGGR_2020.pdf. Accessed 26 Jan 2021
World Health Organization (2014) Gender, women and health: what do we mean by "sex" and "gender"? Available via https://www.legal-tools.org/doc/a33dc3/pdf/. Accessed 4 Nov 2024

Acknowledgements

The authors thank Xiangyang (Shwarn) Cui, Zeshen, Nina Chang, and Stephanie Burri for their help on the project. We also thank Stephanie Burri and Laetitia Burkhard for

proofreading an earlier version of the manuscript. Preparation of this manuscript was supported in part by the Swiss National Science Foundation under grant IZSEZ0_192459, Intramural Research Program of the NIH, NIDA, the Penn Global Research Engagement Fund, and NIH-NIMHD (R01MD018340). The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Author contributions

DP, SCG, LU conceptualized the project. DP wrote the main manuscript text and coordinated the project. TL and AB helped draft the initial version of the manuscript. GS and RY performed the first round of analyses. SR and YMC performed the second round of analyses and prepared all figures. TL and DP created the tables. LU supervised the project. All authors discussed the results, commented on, and extensively reviewed and edited the manuscript.

Competing interests

The authors declare no competing interests.

Ethical approval

This research involved two separate datasets and received appropriate ethical oversight from the Institutional Review Board (IRB) at the University of Pennsylvania (Federal-wide Assurance #00004028). (1) Weibo Data: This dataset was reviewed under the protocol titled *The World Well-Being Project (WWBP) – Weibo* (Protocol #829811) and determined that it met the criteria for exemption under 45 CFR 46.104, Category 4. The exemption was granted by IRB on May 21, 2018. The study involved only secondary analysis of publicly available data. Indirect identifiers or metadata that could pose re-identification risks were excluded from the dataset. (2) Facebook Data: The dataset was reviewed under the protocol titled *Human Flourishing in Social Network (Facebook) Data* (Protocol # 813820) and determined that it met the criteria for exemption under 45 CFR 46.101, Category 4. The exemption was granted by the IRB on June 20, 2011. The data were originally collected by a third party via the myPersonality Facebook application between 2009 and 2011; for the present study, we used data collected in 2010 and 2011. Participants consented to their data being used for academic research at the time of collection. Our research team was not involved in data collection, did not receive any identifying information, and had no means of re-identifying participants. The dataset was provided in fully de-identified form, in accordance with the IRB-approved protocol. As the research team has no available means to re-identify the participants, GDPR regulations did not apply. All members of the research team completed CITI certification in the protection of human subjects, as required by the University of Pennsylvania. Data were stored on secure, institution-controlled servers, subject to annual privacy and security audits. All procedures were conducted in accordance with the Declaration of Helsinki (2013 version) and relevant institutional and U.S. federal regulations and guidelines.

Informed consent

This research involved two separate datasets, each with different consent requirements: (1) Weibo Data: Informed consent was waived for the use of public Weibo data. The Institutional Review Board of the University of Pennsylvania granted an exemption (Protocol #829811, approved May 21, 2018) under 45 CFR 46.104, Category 4, which permits secondary research use of publicly available, non-identifiable data. No private data, indirect identifiers, or metadata enabling re-identification were collected, and no contact with individuals occurred. (2) Facebook Data: The data were originally collected via the *myPersonality* Facebook application, operated by a third party. At the time of participation, users explicitly consented to the academic use of their status updates, demographic information, and psychological assessments. Our research team did not collect the data, was not involved in the consent process, and never received identifiable information. The dataset was provided to us in fully de-identified form, as approved under IRB Protocol #813820 (approved June 20, 2011). Because the data were anonymized and no re-identification was possible, further contact with participants was neither feasible nor required. Data retention and archiving follow the University of Pennsylvania's institutional policies for human subjects research, with secure storage maintained as long as necessary for academic, ethical, and regulatory compliance.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-025-05927-0.

Correspondence and requests for materials should be addressed to Dandan Pang.