Journal of Personality and Social Psychology

Social Media Users Produce More Affect That Supports Cultural Values, but Are More Influenced by Affect That Violates Cultural Values

Tiffany W. Hsu, Yu Niiya, Mike Thelwall, Michael Ko, Brian Knutson, and Jeanne L. Tsai

Online First Publication, September 6, 2021. http://dx.doi.org/10.1037/pspa0000282

CITATION

Hsu, T. W., Niiya, Y., Thelwall, M., Ko, M., Knutson, B., & Tsai, J. L. (2021, September 6). Social Media Users Produce More Affect That Supports Cultural Values, but Are More Influenced by Affect That Violates Cultural Values. *Journal of Personality and Social Psychology*. Advance online publication. http://dx.doi.org/10.1037/pspa0000282

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Personality and Social Psychology:
Attitudes and Social Cognition

ISSN: 0022-3514

https://doi.org/10.1037/pspa0000282

INNOVATIONS IN SOCIAL PSYCHOLOGY

Social Media Users Produce More Affect That Supports Cultural Values, but Are More Influenced by Affect That Violates Cultural Values

Tiffany W. Hsu[1], Yu Niiya[2], Mike Thelwall[3], Michael Ko[1], Brian Knutson[1], and Jeanne L. Tsai[1]
[1] Department of Psychology, Stanford University
[2] Department of Global and Interdisciplinary Studies, Hosei University
[3] School of Mathematics and Computing, University of Wolverhampton

Although social media plays an increasingly important role in communication around the world, social media research has primarily focused on Western users. Thus, little is known about how cultural values shape social media behavior. To examine how cultural affective values might influence social media use, we developed a new sentiment analysis tool that allowed us to compare the affective content of Twitter posts in the United States (55,867 tweets, 1,888 users) and Japan (63,863 tweets, 1,825 users). Consistent with their respective cultural affective values, U.S. users primarily produced positive (vs. negative) posts, whereas Japanese users primarily produced low (vs. high) arousal posts. Contrary to cultural affective values, however, U.S. users were more influenced by changes in others' high arousal negative (e.g., angry) posts, whereas Japanese were more influenced by changes in others' high arousal positive (e.g., excited) posts. These patterns held after controlling for differences in baseline exposure to affective content, and across different topics. Together, these results suggest that across cultures, while social media users primarily produce content that supports their affective values, they are more influenced by content that violates those values. These findings have implications for theories about which affective content spreads on social media, and for applications related to the optimal design and use of social media platforms around the world.

Keywords: culture, emotion, ideal affect, Twitter, contagion

Supplemental materials: https://doi.org/10.1037/pspa0000282.supp

Across cultures, social media platforms have rapidly become a primary channel for communication. Research reveals that people post content on social media for a variety of reasons (Brady et al., 2020; Lin & Utz, 2017; Oh & Syn, 2015). For instance, people may post content that reflects their feelings and values (e.g., users write excited tweets because they feel or want to show excitement). Indeed, an emerging line of research has focused on "affect prevalence," or the types of affective content people produce on social media. This work demonstrates that users in the United States and Western Europe overall tend to produce more positive than negative affective content on social media (e.g., Bazarova et al., 2012; Lin & Utz, 2015; Reinecke & Trepte,

Tiffany W. Hsu https://orcid.org/0000-0003-2525-4261
Yu Niiya https://orcid.org/0000-0002-2733-3724
Mike Thelwall https://orcid.org/0000-0001-6065-205X
Brian Knutson https://orcid.org/0000-0002-7669-426X
Jeanne L. Tsai https://orcid.org/0000-0003-4150-8268

The development of Japanese SentiStrength was supported through National Science Foundation (NSF) Grant 1732963 awarded to Jeanne L. Tsai and Brian Knutson. The authors thank the NSF Social Psychology Program and Stanford HAI Faculty Seed Grant Program for their support. The authors also thank H. Markus, E. Thomas, the Stanford Culture and Emotion Lab, and Culture Collab for their feedback at different stages of the project; G. Suzuku, A. Muto, M. Ruiz, and N. Kikuchi for coding Japanese tweets; and J. Cachia, P. Reyes, D. Ibeling, L. Schlick, L. Murata, and Y. Suezawa for coding tweets for topic content.

The study was not formally preregistered. Data were collected through the public Twitter API (https://dev.twitter.com/overview/api). To comply with the Twitter Developer Agreement and Policy, data cannot be publicly shared. Interested researchers can reproduce the results, however, by following procedures described in online supplementary materials. Custom code and accompanying software for analyzing sentiment of Japanese texts are available at https://github.com/tiffanywhsu/japanese-sentistrength. Custom code for collecting Twitter data, analyzing sentiment of English texts, data processing, and data analyses are available at https://github.com/tiffanywhsu/culture-emotional-contagion.

Correspondence concerning this article should be addressed to Jeanne L. Tsai, Department of Psychology, Stanford University, Building 420, Stanford, CA 94305, United States. Email: jeanne.tsai@stanford.edu

2014). Ironically, this positivity bias may be related to decreased self-esteem among U.S. users, because viewing others' positive posts may lead users to evaluate their own lives more negatively (Vogel et al., 2014).

People may also post content that reflects the affective qualities of what they have just read or viewed, such that their posts reflect the influence of others' posts more than their own feelings and values (e.g., users write angry tweets because they just read another user's angry post). Research on "emotional contagion" demonstrates that people can "catch" emotions from others—often automatically and unconsciously—during face-to-face interactions (Barsade, 2002; Hatfield et al., 1993), and now on social media (Chmiel et al., 2011; Coviello et al., 2014; Ferrara & Yang, 2015; Goldenberg & Gross, 2020; Kramer et al., 2014). Moreover, people seem to catch some types of affect more often than others on social media. For instance, in the United States, users seem to be particularly influenced by others' highly arousing negative affect, such as anger, hate, and outrage (Brady et al., 2017; Brady & Crockett, 2019; Crockett, 2017; Vosoughi et al., 2018; Williams, 2018), resulting in "anger bandwagons" (Williams, 2018) and "viral online shaming" (Crockett, 2017). This is of growing concern because the virality of high arousal negative affective (HAN) content has been associated with the dissemination of fake news and increased political polarization (Crockett, 2017; Vosoughi et al., 2018).

Existing findings, however, have primarily been limited to the United States and other Western countries. As a result, the degree to which the social media transmission of affective content reflects Western cultural values or more general processes remains unclear. For instance, some researchers have argued that a bias toward positive content reflects users' need to present themselves in a socially desirable light (e.g., Bazarova et al., 2012; Lin & Utz, 2015; Reinecke & Trepte, 2014). But maximizing the positive (and minimizing the negative) is more desirable in the United States than in Japan and other East Asian countries (Curhan et al., 2014; Heine et al., 1999; Miyamoto et al., 2010; Sims et al., 2015), raising the possibility that there might be less of a positivity bias in social media posts of users from East Asian countries.

Similarly, although some researchers have argued that HAN states are particularly viral because they signal threat (Kelly et al., 2016), these states also violate the value U.S. culture places on positivity. Because people can only "catch" emotions that they have attended to, it is possible that content that violates cultural values may "hijack" attention (Mu et al., 2015), and therefore have a greater affective impact that leads to increased contagion. Consistent with this idea, Kashima and colleagues observed that although stereotype-consistent information is more prevalent in people's communications and can promote social connection, stereotype-inconsistent information is viewed as more unexpected and surprising (Clark & Kashima, 2007; Simpson & Kashima, 2013). Although cultural affective values differ from stereotypes, a similar process might occur: because high arousal negative states violate the U.S. cultural value of positivity, changes in high arousal negative content in social media may be more unexpected and surprising, and therefore may be particularly contagious in the United States. If this is the case, then in cultures that place less of a value on positivity (e.g., Japan and other East Asian countries), high arousal negative content in social media may be less unexpected and surprising, and therefore less contagious on social media than in the United States.

In addition to clarifying the cultural universality versus specificity of these affective processes on social media, cross-cultural comparisons of social media use might also inform efforts to minimize the harmful effects of social media across the world. For example, efforts to curtail the spread of misinformation in the United States might focus on limiting the spread of misinformation that contains high arousal negative content because it is particularly contagious, whereas similar efforts in other countries might instead focus on misinformation that contains other types of affect that are particularly contagious in those cultures.

Therefore, in this research, we compared the prevalence and contagion of different types of affective content on social media in the United States and Japan. Like the United States, Japan is a modern, industrialized, democratic society with prevalent social media use. Researchers have documented, however, that Japanese value different affective experiences than Americans (e.g., Kitayama et al., 2006; Miyamoto & Ma, 2011; Miyamoto et al., 2017; Ruby et al., 2012; Tsai et al., 2016). Documented cultural differences in affective values allowed us to make distinct and disparate predictions within and between these cultures about which patterns of affect prevalence and contagion might support or violate cultural values. We focused on the valuation of "affective states," or feelings that can be categorized in terms of valence (from positive to negative) and arousal (from low to high; Feldman-Barrett & Russell, 1999; Watson & Tellegen, 1985), because decades of research demonstrate that these two dimensions generalize across cultures and languages (e.g., Kuppens et al., 2006; Yik & Russell, 2003) and because research has demonstrated clear cultural differences in the valuation of specific affective states (Tsai, 2007, 2017; Tsai & Clobert, 2019; Tsai et al., 2006).

The Potential Role of Cultural Values on Social Media Behavior

Decades of research indicate that people from North American (U.S., Canada) versus East Asian cultures (Japan, China, Korea) vary in how much they value different affective experiences (see Tsai & Clobert, 2019, for review). Specifically, because of different models of self and personhood, individuals in the United States aim to maximize positive feelings and minimize negative ones, whereas individuals in many East Asian contexts like Japan desire more moderate feelings, and so aim for a greater balance of positive and negative feelings (e.g., Curhan et al., 2014; Heine et al., 1999; Markus & Kitayama, 2010; Miyamoto et al., 2010; Tsai et al., 2006). Based on affect valuation theory, which states that cultural factors shape the affective states that people value and ideally want to feel even more than the affective states they actually feel (Tsai, 2007, 2017), people in the United States should then value positivity more and negativity less than their East Asian counterparts (Japanese, Chinese, Korean), which has been empirically verified (e.g., Sims et al., 2015). Furthermore, because of different interpersonal goals associated with different models of self, these cultures should also differ in their valuation of high and low arousal positive states (Tsai, Miao, et al, 2007); indeed, U.S. individuals value high arousal positive states (e.g., excitement, enthusiasm) more and low arousal positive states (e.g., calm, peacefulness) less than do East Asian individuals (e.g., Park et al., 2017; Ruby et al., 2012; Tsai, Knutson, & Fung, 2006; Tsai, Miao, et al., 2007, but also see Bencharit et al., 2019).

Previous studies have demonstrated that these cultural differences in ideal affect are reflected in popular media, including children's storybooks, women's magazines, and leaders' official website photos (Tsai, 2007; Tsai, Louie, et al., 2007, Tsai et al., 2016). As cultural products, these forms of media are deliberately created by illustrators, magazine editors, and publicists to reflect dominant cultural values, and these products in turn can shape the values of the people who consume them (Boiger et al., 2013; Kim & Markus, 1999). Like storybooks, magazines, and official photos, Twitter posts and other forms of social media content are also cultural products but are arguably more rapidly and less deliberately constructed. This raises the question of whether cultural differences in ideal affect are also reflected in these newer, emerging types of media. To answer this question, we compared the affective content of U.S. and Japanese users' Twitter posts. We focused on the original posts that users produced, rather than the posts that users shared or reposted (i.e., "retweets") in part because users tend to repost about topics that they do not produce (i.e., originally post) themselves (Macskassy & Michelson, 2011). Therefore, we assumed that users' original posts would reflect their cultural values more closely than would their reposts of others' content. Although we focused on original posts in the article, we explored the content of retweets in supplementary analyses.

Design of the Present Study

To examine the role of cultural values in social media behavior, we collected and analyzed originally produced posts ("tweets") from a sample of United States (n = 1,888 users, 55,867 tweets) and Japanese (n = 1,825 users, 63,863 tweets) users on Twitter.com. This research builds upon the existing literature in several ways. First, we include a sample of non-Western users. Second, while previous research focused on either valence or arousal, we included both, which permitted examination of four different affect types: (a) HAN, (b) low arousal negative affect (LAN), (c) low arousal positive affect (LAP), and (d) high arousal positive affect (HAP). This also allowed us to assess positivity and negativity separately, which better reflects the well-documented statistical independence of positivity and negativity in East Asian contexts (e.g., Grossmann et al., 2016; Sims et al., 2015). Third, we collected posts at multiple time points for each user, so that contagion models could track whether being exposed to different types of affect in others' posts (i.e., the posts of users they are following, or their "follows") was associated with subsequent changes in the affective content of each user's posts. This within-user approach allowed us to control for baseline differences in exposure to affective content and to ensure that our results were not due to between-user confounds such as homophily (i.e., when users follow those who are similar to them), an issue that has limited previous work. Fourth, while previous studies have focused either on prevalence or contagion, we assessed both to examine whether they had similar or different relationships with cultural values. Finally, since most readily available text analysis tools only work for the English language, we developed a sentiment analysis program based on SentiStrength (Thelwall, 2017; Thelwall et al., 2010) which could score short Japanese text in terms of valence (positivity, negativity) and intensity/arousal (ranging from 1 to 5). We built this Japanese version of the SentiStrength program from the ground up, using machine learning based on Japanese research assistants' manually coded labels of a large body of Japanese tweets (program available at https://github.com/tiffanywhsu/japanese-sentistrength; see online supplemental materials, Section S1A and S1B for development details).

In sum, this study addresses limitations of previous work in five important ways: (a) by comparing users from two distinct cultures that differ in their affective values, (b) by distinguishing between low and high arousal positive and negative states, (c) by controlling for baseline differences in exposure and homophily when assessing contagion, (d) by assessing both affect prevalence and contagion, and (e) by developing a tool for analyzing Japanese sentiment in short text.

Hypotheses

We tested two alternative hypotheses regarding the prevalence of affective content. If users overall produce affective content that *supports* their cultural values, then U.S. users should post more positive than negative content, but Japanese users should post more low arousal (i.e., more moderate) than high arousal affective content. In direct comparison, U.S. users should also post more high arousal positive, less low arousal positive content, and less negative content (both high and low arousal) compared with Japanese users. Alternatively, if users overall produce content that violates their cultural values, then the opposite patterns should emerge both within and between cultures.

We tested these same hypotheses with respect to the contagion of affective content. If users are more influenced by others' affective content that supports their cultural values, then U.S. users' posts should be more influenced by changes in exposure to others' positive content than others' negative content. Further, Japanese users' posts should be more influenced by changes in exposure to others' low arousal content than others' high arousal content. In direct comparison, U.S. users should be more influenced by changes in exposure to high arousal positive content, and less influenced by changes in exposure to low arousal positive content and negative (both high and low arousal) content than Japanese users. Again, however, if users are more influenced by changes in content that violates their cultural values, then opposite patterns should emerge both within and between cultures.

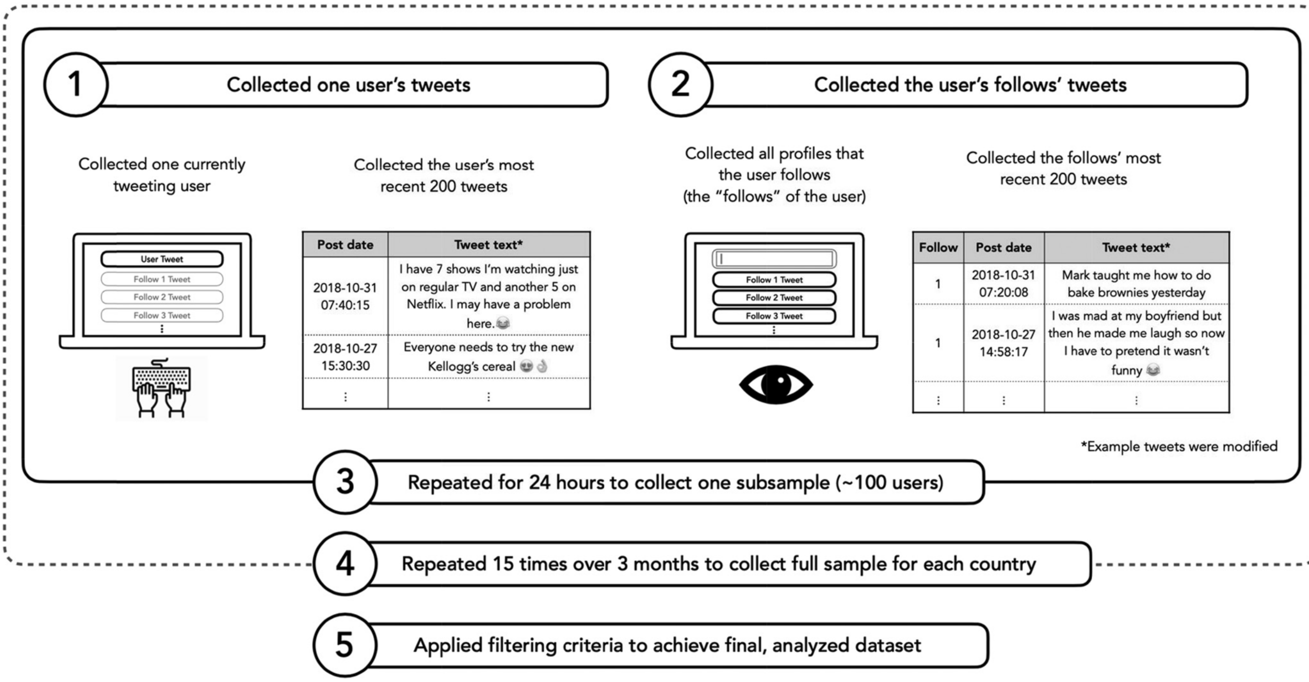
If cultural values do not influence affect prevalence or contagion, then Japanese and U.S. users should both produce more positive than negative content, and be primarily influenced by changes in others' high arousal negative affect, as documented in previous research on Western users (Bazarova et al., 2012; Crockett, 2017; Lin & Utz, 2015; Reinecke & Trepte, 2014).

Method

Data Collection

Using the Python package *tweepy* and the Twitter Application Programming Interface (API), we collected (a) tweets posted by a set of users located in the United States and Japan, defined as the latitude/longitude geographical boundary of the two countries as set by Twitter, and (b) tweets posted by the profiles that users followed to assess users' exposure to others' affective content (see Figure 1). Because the Twitter API lacks the functionality to collect a random sample of typical users, we collected subsamples of users posting at various times and days, over a course of three

Figure 1
Process Diagram Depicting Collection of Users' Tweets and Users' Follows' Tweets



months, achieving a final sample of users as close to the typical Twitter user as possible. For each subsample of users, we used the Twitter Standard Streaming API to collect one random tweet posted at a time, extracted the user ID from the tweet, and included the user in our sample if (a) the language of the user's Twitter platform was set to English for the U.S. users or to Japanese for the Japanese users; (b) the language of the tweet was detected by Twitter as English or Japanese; and (c) the user passed a bot check using the package *botometer* (bot scores ranged continuously from 0 to 5, with 0 being most human-like and 5 being most bot-like, and we admitted only users with bot scores of 1 or less; http://botometer.iuni.iu.edu).

We then used the Twitter Standard Search API to collect the user's most recent 200 tweets, due to time constraints imposed by the Twitter API rate limits (for additional details on the data collection rationale, see online supplemental materials, Section S2). Consistent with previous studies on emotional contagion on Twitter (e.g., Ferrara & Yang, 2015), these original tweets included the tweets users posted on their timelines, quote tweets (without the retweet component), and replies, all of which could be influenced by users' previous exposure to others' posts. To approximate recent exposure in assessing contagion, we then collected the entire set of profiles that the targeted user followed (the "follows") and collected the follows' most recent 200 tweets.

For each subsample, we repeated the above procedure for 24 continuous hours to collect different users who were posting at different times of the day. We then collected different subsamples over a span of three months (early October 2018 to January 2019), varying the day of week of collection, until we reached an approximate total of 4,000 users, based on Ferrara and Yang (2015). This sample size was large enough to provide sufficient power to mitigate against the noisiness of real-world data, without overwhelming the rate limits of the Twitter API.

In total, we collected 523,810 tweets (219,752 from the United States, 304,058 from Japan) from 4,056 users (2,045 from the United States and 2,011 from Japan). For these users (without accounting for duplicates across users), there were 3,437,324 follows (2,035,768 for the U.S. users; 1,401,556 for the Japanese users), from whom we collected a total of 455,545,112 tweets (272,299,371 from the United States; 183,245,741 from Japan).

After collection, we applied several criteria to filter tweets before analysis. Based on the criteria described in Ferrara and Yang (2015), we excluded user tweets that had fewer than 20 corresponding follows' tweets used in calculating exposure. Because users' follows could change over time (i.e., users could have followed and unfollowed profiles) but we could only collect the set of users' follows at the time of collection, we also excluded user tweets (and their corresponding follows tweets) that were posted more than a week before the collection date to better ensure that the set of follows were accurate. U.S. and Japanese users posted an average of 49.4 tweets (SD = 47.6) and 73.5 tweets (SD = 60.4), respectively, within a week before the collection dates; therefore, we analyzed only the most recent 50 tweets to reduce the range of tweets examined, and to equate these ranges for U.S. and Japanese users. These procedures resulted in a set of 1889 U.S. users (55,917 user tweets) and 1836 Japanese users (64,390 user tweets). Out of that group, one U.S. user and 11 Japanese users were duplicated (i.e., had their tweets collected twice) based on Twitter user ID matching; we removed the duplicate tweets for these users before analyses.

Thus, the final sample that we analyzed was comprised of 1,888 U.S. users (55,867 user tweets) and 1,825 Japanese users (63,863 user tweets); U.S. users had an average of 29.6 tweets (*SD* = 19.0), and Japanese users had an average of 35.0 tweets (*SD* = 18.4; for histograms, see online supplemental materials, Section S3A, Figure S2). To calculate exposure corresponding to each user tweet, we further filtered follows' tweets so that we only included those that were posted at most an hour before each user tweet, per criteria set in Ferrara and Yang (2015). Users' exposure to different types of affective content was based on an average of 316.73 tweets (*SD* = 812.46) from 101.36 follows (*SD* = 199.06) for U.S. users, and 207.53 tweets (*SD* = 330.06) from 78.82 follows (*SD* = 100.00) for Japanese users.

Sentiment Analysis and Categorization

The SentiStrength algorithm used to label the affective content of posts (Thelwall et al., 2010) relies on a dictionary set that includes terms labeled by valence and intensity (e.g., "anger" = negativity 4; "calm" = positivity 2), as well as semantically relevant terms such as booster words (e.g., "extremely"), negating words (e.g., "couldn't"), question words (e.g., "why"), emojis (for example, ":("; see updated list in the online supplemental materials, Section S1C), slang words (for example, "lol"), and domain-specific terms (for example, "must watch" in the context of film). The program then optimizes the term labels using machine learning trained on a set of human-labeled social media web texts (Thelwall, 2017).

We chose SentiStrength because (a) it was developed to detect the sentiment of short social media web text samples (e.g., Twitter posts) and has been used for this purpose in previous research (Ferrara & Yang, 2015); (b) it provides separate scores for positivity and negativity, which was critical for this study, given cultural differences in the statistical independence of positivity and negativity between East Asian and Western samples (e.g., Grossmann et al., 2016; Sims et al., 2015; see online supplemental materials, Section S4A for data on "mixed" tweets), and (c) it codes intensity/arousal for each positivity and negativity code, allowing us to examine whether cultural differences in the valuation of specific states defined in terms of valence and arousal were reflected in the affective content of users' posts.

Different Affect Types

Although intensity and arousal are not theoretically identical, they are often correlated in self-report (Kuppens et al., 2013); and are coded similarly in SentiStrength. Based on these codes, we categorized tweets as follows: LAP tweets were those that received a SentiStrength positivity score of 2; HAP tweets were those that received SentiStrength positivity scores of 3, 4, or 5; LAN tweets were those that received SentiStrength negativity scores of 2; and HAN tweets were those that received SentiStrength negativity scores of 3, 4, and 5 (see Table 1 for examples of coded tweets). We used 3 and above to indicate high arousal because words psychometrically associated with "high arousal" (e.g., "excitement") were assigned a score of 3 by SentiStrength. "Neutral [NEU] or uncodable" tweets were those that received both positivity and negativity scores of 1 indicating no positivity or negativity, respectively. Because we were primarily focused on affect prevalence and contagion, and because the overall pattern of results remained the same when neutral tweets were included in our analyses, we do not present the neutral tweets here, but interested readers should see online supplemental materials, Section S3G.

Development of Japanese SentiStrength

Prior to this study, SentiStrength did not have a Japanese version, and no readily available tool existed to analyze the valence and intensity of short Japanese text. Thus, we developed a version of SentiStrength for Japanese by (a) compiling a set of human-rated sentiment dictionary terms in Japanese, (b) developing program capabilities to accommodate particular characteristics of the Japanese language, and (c) optimizing the program based on a training set of Japanese tweets coded for affective content by Japanese native speakers living in Japan (for details on development procedures, see online supplemental materials, Section S1A).

To assess the performance of this Japanese version of SentiStrength, we (a) applied the program to a test set of human-rated Japanese tweets, and (b) validated the findings from Japanese SentiStrength by comparing them to findings from the human raters (for details on performance procedures, see online supplemental materials Section S1B).

Performance of Japanese SentiStrength

We assessed accuracy with metrics used in the development of the English version of SentiStrength (Thelwall et al., 2010); and with metrics comparing the two SentiStrengths. We also compared the accuracy scores of Japanese SentiStrength with other current state-of-the-art models (Barnes et al., 2017). Because the SentiStrength scores were grouped into affect types for this study, we report the performance metrics for each affect group: for positivity, we conducted separate analyses for three groups: (a) positivity 1, (b) positivity 2 [LAP], and (c) positivity 3,4,5 [HAP]; for negativity, we conducted separate analyses for three groups: (a) negativity 1, (b) negativity 2 [LAN], and (c) negativity 3,4,5 [HAN]. Ceiling accuracy was the average human interrater agreement, which was 63.8% for classifying the positive types, and 69.1% for classifying the negative types among Japanese raters; these accuracies were comparable to the average human interrater agreement in classifying raw affect scores for English SentiStrength (Thelwall et al., 2010).

Table 1
Examples of Categorized Tweets (With Identifying Content Removed)

Affect type	Example tweet
LAP (pos = 2, neg = 1)	I like your hair
HAP (pos = 3, neg = 1)	The cutest pictures are from Kindergarten graduation!😍❤️
LAN (pos = 1, neg = 2)	Another week of exams then I'm sorta free 🏀😒
HAN (pos = 1, neg = 5)	Roze Rizee is a TERRIBLE singer and a heinous person!

Note. LAP = low arousal positive affect; HAP = high arousal positive affect; LAN = low arousal negative affect; HAN = high arousal negative affect.

The overall accuracy of Japanese SentiStrength was 53.8% for classifying the positive types ($p < .001$ derived from permutation testing) and 52.5% for classifying the negative types ($p < .001$ derived from permutation testing). These accuracies scores were at least 10% lower than the ceiling accuracies described above (positive: 53.8% versus 63.8%; negative: 52.5% versus 69.1%), which is not surprising, given the considerable difficulty of classifying valence and arousal (versus valence only; Barnes et al., 2017). Notably, the accuracy scores of Japanese SentiStrength are higher than 45.6%, which is the highest accuracy score obtained by current state-of-the-art models trained and tested on the Stanford Sentiment Treebank (SST; Socher et al., 2013), an English-language dataset labeled with five levels of sentiment from 'strongly negative' to 'strongly positive' (Barnes et al., 2017). SST was a relevant comparison because like our program, it distinguished between different types of positive and negative content. Thus, Japanese SentiStrength—like English SentiStrength—-outperformed these state-of-the-art models.

Because the positivity and negativity groups were imbalanced (about half of the tweets received scores of 1 for both positivity [47.5%] and negativity [57.3%]), we calculated weighted F1 scores on classification of these groups to assess precision and recall. We computed weighted F1 scores of one randomly selected human rater's ratings, compared to the average of the other three human raters' ratings to obtain a "ceiling F1 score" for the positive and the negative groupings. We found "ceiling" weighted F1 scores of .617 for the positive groupings and .603 for the negative groupings. The F1 scores for Japanese SentiStrength were .506 for the positive groupings and .489 for the negative groupings. Thus, the F1 scores for Japanese SentiStrength were about .1 lower than the ceiling F1 scores (.51 versus .62, .49 vs .60). Because the negative groupings had an F1 score of less than .5, we conducted additional analyses to further demonstrate the validity of the program and our results in comparison to the human ratings.

Specifically, we assessed the extent to which errors in Japanese SentiStrength classification reflected systematic differences in how Japanese human raters classified affect types. We found that the rates at which Japanese SentiStrength classified or misclassified tweets was highly correlated with the rates at which one human rater agreed or disagreed with the average of the other human raters: $r(7) = .965$, $p = .000$ for positive groupings, and $r(7) = .875$, $p = .002$ for negative groupings. These numbers show that errors in Japanese SentiStrength classification might reflect the nature of distinguishing between these affect categories among Japanese users (for further details on this analysis, see online supplemental materials, Section S1B, Table S4). Thus, to ensure that our main results were not due to inherent artifacts in Japanese SentiStrength but reflected the actual content of Japanese posts, we also conducted the same prevalence analyses using the human ratings of the 3,481 tweets used in Japanese SentiStrength development. The overall pattern of results based on the human ratings was similar to the pattern of results based on the Japanese SentiStrength ratings (as described below and reported in online supplemental materials, Section S4B).

Finally, we compared the performance of English SentiStrength and Japanese SentiStrength to ensure that the study results were not due to differential sensitivities between the two programs in classifying each affect type (online supplemental materials, Section S1B, Table S5). For each valence, we calculated the percentage of tweets that were correctly scored by SentiStrength as one group and incorrectly scored as the two other groups. This generated a 3×3 matrix of percentages for each valence. We calculated these matrices for both English SentiStrength and Japanese SentiStrength (using a separate set of previously human-coded 4,218 random English tweets). Given that we were not specifically interested in neutral tweets (i.e., scores of positivity = 1 and negativity = 1) in our study, we removed from the matrices the cells corresponding to tweets that were categorized by the two programs as neutral. Finally, we compared the positivity matrices between the two programs by correlating the matrices and then did the same for the negativity matrices. The matrices were moderately correlated at $r(4) = .61$ for positivity and highly correlated at $r(4) = .80$ for negativity, suggesting that while English and Japanese SentiStrength programs showed similar degrees of sensitivity for negativity, they showed slightly different degrees of sensitivity for positivity. Because the confusion matrices for Japanese SentiStrength and Japanese human raters were highly correlated, however, it is possible that these differences in sensitivity might reflect the nature of Japanese versus English linguistic expression of positivity, or the detection of positivity in Japanese versus English text, rather than a limitation of Japanese SentiStrength per se (see Discussion).

In sum, Japanese SentiStrength performed slightly worse than the ceiling metrics of human interrater agreement, but its errors likely reflect how Japanese human raters distinguish between affect scores, and its accuracies were comparable to current state-of-the-art models trained on English data sets with similar sentiment labels. Moreover, the correlations among human raters and SentiStrength were significantly positive, indicating that across the tweets, human raters and SentiStrength rated less intense tweets as less intense, and more intense tweets as more intense (see Thelwall et al., 2010; and online supplemental materials, Section S1B, Tables S1 and S3 for these metrics). By grouping the tweets into the four affect types, we could ensure that the high arousal set of tweets would on average still be higher in intensity than the low arousal set of tweets. Japanese SentiStrength showed similar sensitivity in classifying negativity but slightly different sensitivity in classifying positivity compared to English SentiStrength. Despite these limitations, we believe that SentiStrength—in Japanese and English—still provides a valid assessment of linguistic expressions of sentiment expressed in short text by identifying sentiment-related words, phrases, and emojis, similar to programs such as the Linguistic Inquiry and Word Count Program (Pennebaker et al., 2015); and those used by Kramer et al. (2014).

Data Analyses and Results

Affect Prevalence: Do Social Media Users Produce Affective Content That Supports or Violates Their Cultural Values?

We first addressed whether Twitter users overall post affective content that supports or violates their cultural affective values. To examine the prevalence of different types of affect in users' tweets, we calculated the overall percentage of tweets categorized as HAP, LAP, HAN, and LAN for each user. To compare percentages between affect types within culture, we fitted mixed linear regression models using affect type to predict percentage with random intercept of user. To compare percentages of each affect type between cultures, we fitted a mixed linear regression model using culture (0 = Japan, 1 = U.S.) to predict percentage with random intercept of user. Because of the large sample sizes, most estimates were significant ($p < .01$); therefore, we also used Cohen's h to indicate the size of the effects. We first averaged the percentages across users for each culture to obtain the overall percentages of user tweets categorized as HAP, LAP, HAN, and LAN for each user. Cohen's h between two percentages (p1 and p2) was calculated as $h = 2 \times abs\left(arcsin\left(\sqrt{\frac{p1}{100}}\right) - \ arcsin\ arcsin\ \left(\sqrt{\frac{p2}{100}}\right)\right)$. Effect sizes of less than .2 were considered small; effect sizes between .2 to .5 were considered medium, and effect sizes greater than .5 were considered large based on Cohen's rule-of-thumb guidelines (Cohen, 1988).

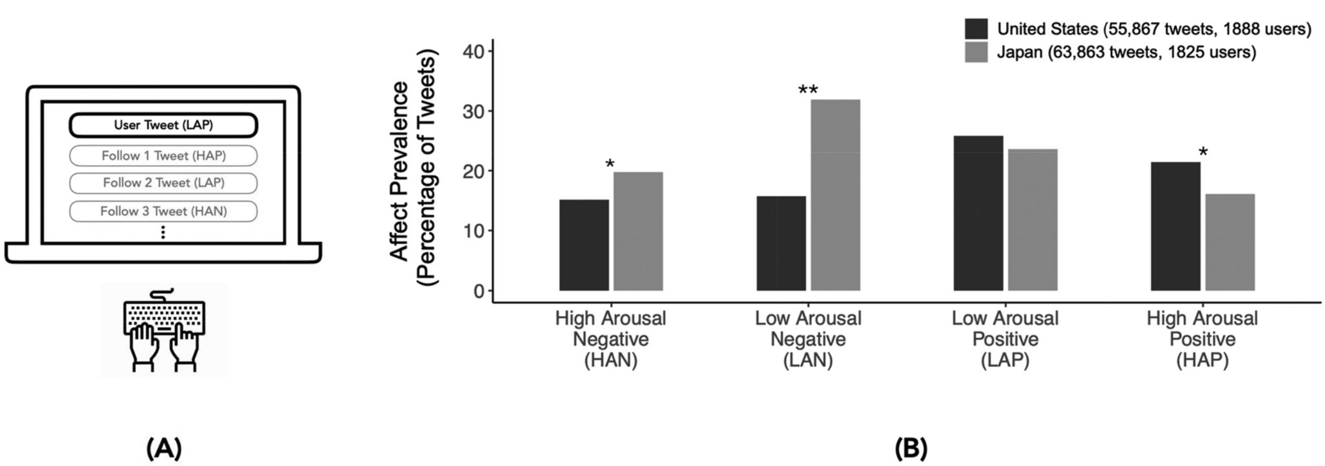
Within-Culture Comparisons

United States. Consistent with U.S. affective values, U.S. users posted more positive than negative content (see Figure 2B, black bars; positive: 47.28% of tweets overall, broken down into 21.45% HAP and 25.84% LAP; negative: 30.86% of tweets overall, broken down into 15.14% HAN and 15.72% LAN), $b = 16.43$, $SE = .671$, $t = 24.48$, $p < .001$, $h = .34$), replicating previously-documented patterns (Reinecke & Trepte, 2014). Pairwise comparisons across all affect types specifically revealed that U.S. users posted more low and high arousal positive content than low and high arousal negative content (ps < .001, hs range from .15 to .27). U.S. users also posted more low arousal positive than high arousal positive content, although this effect was small ($p < .001$, $h = .10$), and they did not differ in their posting of high arousal negative versus low arousal negative content ($p = .203$, $h = .02$: see online supplemental materials, Section S3B for pairwise comparison statistics).

Japan. Analyses of Japanese tweets, however, revealed a different pattern (see Figure 2B, gray bars). Consistent with Japanese affective values, Japanese users overall posted more low arousal content than high arousal content (low arousal: 55.53% of tweets overall, broken down into 31.90% LAN and 23.63% LAP; high arousal: 35.84% of tweets overall, broken down into 19.75% HAN and 16.09% HAP), $b = 19.69$, $SE = .828$, $t = 23.77$, $p < .001$, $h = .40$). Specific pairwise comparisons also revealed that Japanese users posted more low arousal negative and low arousal positive content than high arousal negative and high arousal positive content (ps < .001, hs range from .09 to .37). Japanese users also posted more low arousal negative than low arousal positive content ($p < .001$, $h = .19$), and more high arousal negative than high arousal positive content ($p < .001$, $h = .10$). The greater prevalence of low arousal negative and low arousal positive compared with high arousal negative and high arousal positive content is consistent with the notion that Japanese would post more moderate and balanced affective content (see online supplemental materials, Section S3B for pairwise comparison statistics).

Interestingly, Japanese users posted overall more negative content than positive content, although the effect sizes were small. This was a pattern we did not predict. Based on our review of the tweets, it appeared that some of the negative content had positive connotations (e.g., ありがたきアル中!退屈だから飲むのであります!, translated as "Grateful to be alcoholic! I drink because I am bored!"). One possible explanation is that some of the negative content may have been intended to be self-deprecating and self-effacing, which are desirable in Japan (Tsukawaki et al., 2011) because they signal the cultural valuation of self-improvement (Heine et al., 1999). Indeed, self-deprecating humor is associated with better mental health and more positive evaluation by others in Japan (Tsukawaki et al., 2011; Yoshida et al., 1982). The greater prevalence of negative content in Japanese tweets may also reflect a desire to elicit sympathy in others (Kitayama & Markus, 2000). The development of more nuanced

Figure 2
Cultural Variation in Affect Prevalence



Note. A: User tweets coded for prevalence of each affect type (% user tweets with specific affect type); B: Affect prevalence by cultural group. Between-culture effect sizes = ** Cohen's $h > .2$ (medium), * $.2 >$ Cohen's $h > .1$ (small). HAN = high arousal negative affect; LAN = low arousal negative affect; LAP = low arousal positive affect; HAP = high arousal positive affect.

coding systems would allow us to examine these hypotheses in the future.

Between-Culture Comparisons

Consistent with cultural differences in affective values, U.S. users posted more high arousal positive content (U.S. users: 21.45%, Japanese users: 16.09%, $b = 5.35$, $SE = .559$, $t = 9.57$, $p < .001$, $h = .14$) and less overall negative content (overall negative: U.S. users: 30.86%, Japanese users: 51.66%, $b = -20.80$, $SE = .710$, $t = -29.30$, $p < .001$, $h = .43$; broken down into HAN: U.S. users: 15.14%, Japanese users: 19.75%, $b = -4.62$, $SE = .501$, $t = -9.22$, $p < .001$, $h = .12$; LAN: U.S. users: 15.72%, Japanese users: 31.90%, $b = -16.18$, $SE = .599$, $t = -27.01$, $p < .001$, $h = .39$) than did Japanese users. Contrary to cultural differences in affective values, however, U.S. users posted more low arousal positive content than did Japanese users, although the size of this effect was very small relative to the other cultural differences ($b = 2.20$, $SE = .548$, $t = 4.02$, $p < .001$, $h = .05$). Although the size of this effect was small, it may reflect more recently-observed increases in the valuation of low arousal positive affect among European Americans (Bencharit et al., 2019; Tsai et al., 2019).

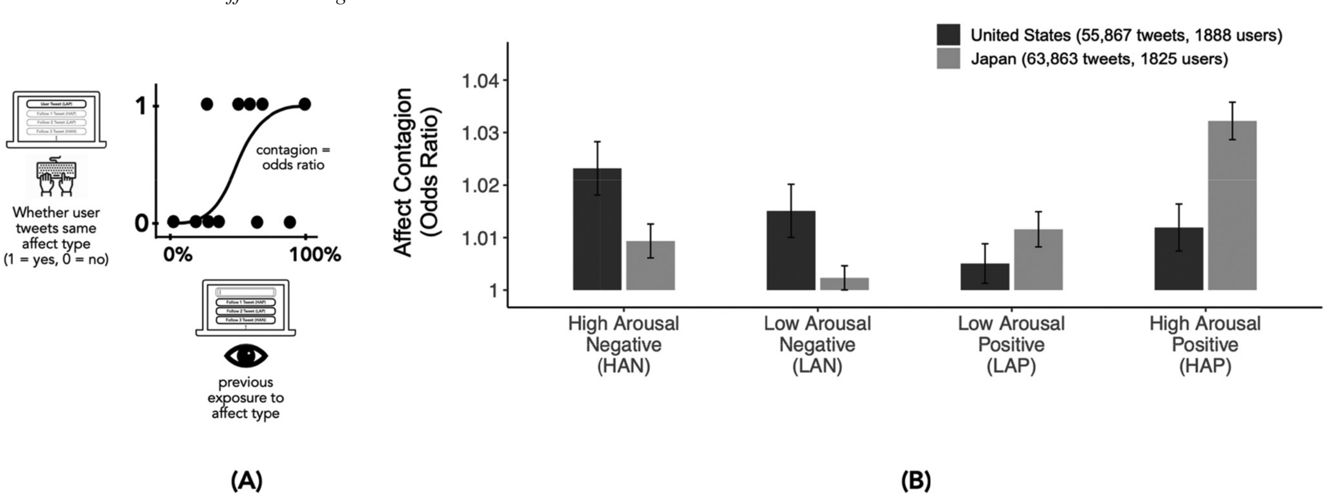
In sum, consistent with cultural values, U.S. users posted more positive than negative content, whereas Japanese users posted more low arousal than high arousal content. Moreover, when directly compared, U.S. users posted more high arousal positive and less negative (both high and low arousal) content than did Japanese users. These medium-sized effects are consistent with cultural affective values. Although U.S. users posted more low arousal positive content than did Japanese users, the size of this effect is small. Therefore, the largest effects are consistent with the notion that both within and between cultures, people tend to post affective content that *supports* their cultural values.

Affect Contagion: Are People More Influenced by Affective Content That Supports or Violates Their Cultural Values?

After determining which affective content U.S. and Japanese users primarily produced in their original posts, we examined which type of affective content they were most "influenced" by in others' posts. Like other observational studies of contagion (Ferrara & Yang, 2015); our data were correlational, and therefore, we could only approximate causal influence by focusing on follows' tweets that were posted before each user tweet. If users are influenced by affective content that *supports* their cultural values, then positive content should be more contagious than negative content for U.S. users, and low arousal content should be more contagious than high arousal content for Japanese users. Based on the prevalence findings, U.S. users should be more influenced by high arousal positive content and less by negative content than Japanese users. However, if users are instead more influenced by content that *violates* their cultural values, then the opposite patterns should hold.

To test these predictions, we first quantified each user's exposure to all four affect types prior to producing original posts. This was calculated as the percentage of tweets posted by the user's follows within one hour before the user posted each tweet (Ferrara & Yang, 2015). For example, for tweet *i* posted by user *j*, if 10% of tweets posted by *j*'s follows an hour before *i* was posted contained HAP, then *j*'s exposure to HAP prior to posting tweet *i* was 10%. Then, for each culture, we fitted a single multinomial multivariate logistic regression model that predicted whether each tweet posted by each user contained each of the four affect types (HAP, LAP, HAN, LAN), using exposure to all four affect types as predictors (Figure 3A; for more details, see online supplemental materials, Section S3E). We quantified the degree to which users were "influenced" by their follows' posts for each specific affect type as the odds ratio derived from this model. These odds ratios captured the

Figure 3
Cultural Variation in Affect Contagion



Note. A: Affect contagion coding as change in likelihood of tweeting same affect type, given a 1% change in previous exposure (odds ratio); B: Affect contagion by cultural group. Error bars represent 95% confidence interval. HAN = high arousal negative affect; LAN = low arousal negative affect; LAP = low arousal positive affect; HAP = high arousal positive affect.

extent to which a 1% change in users' exposure to the affect type changed the odds that the subsequent user tweet contained a specific affect type; for example, an odds ratio of 1.05 for HAN would mean that a 1% increase (or decrease) in users' exposure to HAN increased (or decreased) the likelihood that the user would subsequently produce a tweet with HAN by 5%.

Critically, the model uses random intercepts of user to control for between-user differences in average exposure to each affect type. These random intercepts also address the commonly-observed confound of homophily (i.e., users tend to follow those who are similar to them; McPherson et al., 2001), ensuring that the observed odds ratios captured how much changes in exposure were associated with subsequent posting within users. Although assessing fixed effects among users would be ideal, for privacy reasons, the Twitter API does not release information about users needed for such analyses. The model also included random intercepts of post date to control for another common confound of "exogenous shocks" (i.e., common events that both users and their follows concurrently experience that might trigger similar emotional responses).

Thus, an affect type with an odds ratio greater than 1 indicates that a 1% change in exposure to that affect type changed the likelihood of the user producing a post with a particular affect type in the *same* direction (i.e., an increase in exposure increased the likelihood, and a decrease in exposure decreased the likelihood), suggesting that the affect type was "contagious." An affect type with an odds ratio equal to 1 would mean that a 1% increase in exposure had no effect on the user's likelihood of generating a post with that affect type, suggesting that the affect type was "not contagious." Finally, an affect type with an odds ratio less than 1 indicates that a 1% change in exposure to that affect type changed the likelihood of the user producing a post with that same affect type in the *opposite* direction (i.e., an increase in exposure *decreased* the likelihood of producing a post, or a decrease in exposure *increased* the likelihood of producing a post).

The model was fitted separately for U.S. and Japanese users. To compare odds ratios between affect types, we conducted chi-squared tests using the linearHypothesis function from the R *car* package. All *p*-values were one-sided (as per chi-squared test conventions; for model details, see online supplemental materials, Section S3E).

Although the model included congruous pairs (e.g., exposure to HAP predicting production of HAP), it also included incongruous pairs (e.g., exposure to HAN predicting production of HAP). However, as shown in the full output (online supplemental materials, Section S3E, Table S10a), for most affect types, the effects of congruous pairs were stronger than the effects of incongruous pairs. In other words, exposure to an affect type was most influential in changing the likelihood of the user posting that same (congruous) affect type. Therefore, we focus on congruous pairs here (but see online supplemental materials, Section S3E, Table S10a for results with incongruous pairs).

Analyses revealed that all four affect types were contagious in both the United States and Japan (odds ratios were significantly greater than 1, $ps < .05$), supporting previous findings of emotion contagion on social media (Ferrara & Yang, 2015; Kramer et al., 2014). In other words, when people are exposed to increases (or decreases) in affective content based on their follows' posts, they are in general more (or less) likely to produce similar affective content. Within each culture, however, the degree of contagion also varied by affect type (see Figure 3B; see online supplemental materials, Section S3E, Table S9 for full model outputs).

Within-Culture Comparisons

United States. Among U.S. users (Figure 3B, black bars), high arousal negative content influenced users more than the other three affect types. Given a 1% change in exposure, the likelihood of U.S. users producing HAN in their subsequent original posts changed by 2.3%, compared to 1.5% for LAN, .5% for LAP, and 1.2% for HAP: HAN $OR = 1.023$, 95% CI [1.018, 1.028], LAN $OR = 1.015$, 95% CI [1.010, 1.020]; LAP $OR = 1.005$, 95% CI [1.001, 1.009]; HAP $OR = 1.012$, 95% CI [1.007, 1.016]); HAN versus LAN $\chi^2(1) = 4.93$, $p = .026$; HAN versus LAP $\chi^2(1) = 31.87$, $p < .001$; HAN versus HAP, $\chi^2(1) = 10.64$, $p = .001$. U.S. users were least influenced by changes in LAP in their follows' posts: LAP versus HAN, $\chi^2(1) = 31.87$, $p < .001$; LAP versus LAN, $\chi^2(1) = 9.74$, $p = .002$; LAP versus HAP, $\chi^2(1) = 5.26$, $p = .022$. There was no difference in how influenced U.S. users were by changes in LAN versus HAP content in their follows' posts, $\chi^2(1) = .85$, $p = .357$.

Thus, U.S. users were most influenced by changes in exposure to high arousal negative content, which violates the U.S. emphasis on maximizing the positive and minimizing the negative. These results corroborate past accounts of the particular virality of high arousal negative affective content observed in English-speaking social media (Brady et al., 2017; Brady & Crockett, 2019; Crockett, 2017; Vosoughi et al., 2018; Williams, 2018). To put these contagion effects in the context of real-world changes in exposure to affect, we calculated the average absolute change in exposure (i.e., difference in exposure from one tweet to the next) across users to examine the average change in likelihood of users posting certain affect types from one tweet to the next. For U.S. users, the average change in exposure across the four affect types was 3.43% (3.11% for HAN, 3.21% for LAN, 3.85% for LAP, and 3.54% for HAP). Given a 3.43% increase in exposure, U.S. users were 8.2% more likely to post a tweet containing HAN, compared to 5.3% for LAN, 1.7% for LAP, and 4.1% for HAP.

Japan. Japanese users (Figure 3B, gray bars), however, were most influenced by changes in the high arousal positive content of their follows compared to the other three affect types. Given a 1% change in previous exposure, the likelihood of users producing HAP in their original posts increased by 3.2%, compared to .9% for HAN, .2% for LAN, and 1.2% for LAP: HAP $OR = 1.032$, 95% CI [1.029, 1.036], HAN $OR = 1.009$, 95% CI [1.006, 1.013]; LAN $OR = 1.002$, 95% CI [1.000, 1.005]; LAP $OR = 1.012$, 95% CI [1.008, 1.015]; HAP versus HAN, $\chi^2(1) = 86.41$, $p < .001$; HAP versus LAN, $\chi^2(1) = 194.22$, $p < .001$; HAP versus LAP, $\chi^2(1) = 69.08$, $p < .001$. In contrast, Japanese users were the least influenced by changes in follows' LAN (LAN versus HAN $\chi^2(1) = 11.58$, $p < .001$; LAN versus LAP $\chi^2(1) = 19.91$, $p < .001$; LAN versus HAP $\chi^2(1) = 194.22$, $p < .001$), and there were no differences in how influenced Japanese users were by changes in HAN versus LAP content in their follows' tweets, $\chi^2(1) = .877$, $p = .349$.

Thus, in Japan, users were most influenced by others' high arousal positive content, which violates the Japanese emphasis on low arousal and balanced affect. Again, to put these contagion effects in terms of real-world changes in exposure, we calculated the average absolute difference in exposure across the four affect types for Japanese users, which was found to be 3.68% (4.14% for HAN, 3.15% for LAN, 3.77% for LAP, and 3.66% for HAP). Given a 3.68% change in exposure to the specific affect types, Japanese users were 12.4% more likely to post a tweet containing HAP, compared to 3.5% for HAN, .9% for LAN, and 4.3% for LAP.

Together with the U.S. findings, these results suggest that users are most likely to be influenced by others' posts when those posts contain affective content that *violates* cultural values.

Between-Culture Comparisons

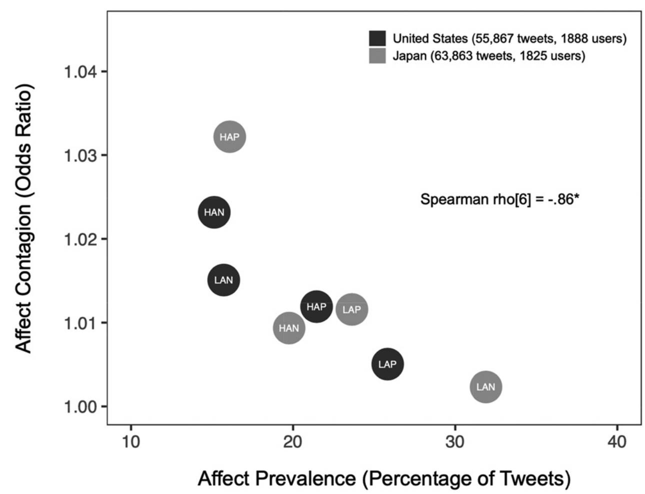
To formally test for cultural differences in which types of affect most influenced users, we fitted a model similar to the contagion model, with an additional dummy variable for culture coded as 1 for U.S. users and 0 for Japanese users; thus, U.S. users were more influenced by affect types if odds ratios were > 1, and Japanese users were more influenced by affect types if odds ratios were < 1 (for more details, see online supplemental materials, Section S3E). Analyses revealed that U.S. users were more influenced by changes in others' negative states (high and low arousal) than were Japanese users (HAN: *OR* = 1.012, CI [1.006, 1.018]; LAN: *OR* = 1.008, 95% CI [1.003, .1.014]), whereas Japanese users were more influenced by changes in others' positive states (high and low arousal) than were U.S. users (HAP: *OR* = .976, 95% CI [.971, .982]; LAP: *OR* = .990, 95% CI [.985, .995]). These findings provide further evidence that users (particularly those from the US) were more likely to be influenced by affect types that *violated* their cultural values.

In sum, in the United States, users were most influenced by changes in others' high arousal negative content, whereas in Japan, users were most influenced by changes in others' high arousal positive content. Moreover, in direct comparison, U.S. users were more influenced by changes in others' negative content than Japanese users, whereas Japanese users were more influenced by changes in others' positive content than U.S. users. These findings suggest that within and between cultures, people are more likely to be influenced by affective content that violates their cultural values. Importantly, these findings were based on original posts. Specifically, users produced original tweets that reflected the previous affective content of their follows' tweets—especially when that affective content violated their cultural values. Because these models controlled for users' baseline exposure and for date of posting, this pattern of results could not be attributed to differences in exposure to affective content due to similar follows or similar exogenous events (for additional analyses on the effects of date, see online supplemental materials, Section S4E).

Together, these findings suggest that while cultural values appear to shape the affective content users produce as well as the type of affective content they are influenced by, they do so in different ways. Although users are more likely to produce affective content that supports their cultural values, they are more likely to be influenced by content that violates those cultural values, suggesting a negative association between the two. To specifically test whether this was the case, we correlated the prevalence (percentage of overall original posts of each affect type) and contagion metrics (the odds ratio of each affect type). To maximize sample size, we collapsed across cultural groups. Across both cultural groups, the more users produced a particular affect type, the less influenced they were by changes in that type of affect in others' posts (Spearman rho [6] = −.86, *p* = .011; see Figure 4). This negative association held within cultural groups as well, although the correlations were not significant, since they were based on fewer data points (US: Spearman rho [2] = −1.00, *p* = .083, Japanese: Spearman rho [2] = −.80, *p* = .333). We also conducted these analyses at the individual user level and observed similar results (see online supplemental materials, Section S3F).

Figure 4

Association Between Prevalence of Affect That Users Produce and the Degree to Which Users Were Influenced by Others' Affect in the United States and Japan



Note. HAP = high arousal positive affect; HAN = high arousal negative affect; LAP = low arousal positive affect; LAN = low arousal negative affect. * $p < .05$.

Ruling Out Topic Content Confounds

One possible alternative explanation for the observed cultural differences is that U.S. and Japanese users discussed different topics in their tweets. Collection of data across a three-month period already decreased the possibility of confounds related to specific events. However, to further rule out a content-based account, four trained research assistants (two American, two Japanese) coded the topics (personal matters, professional matters, entertainment, social commentary, and politics) of 3,500 randomly selected U.S. tweets and 3500 randomly selected Japanese tweets from the above data set. Among this subset of coded original tweets, the majority of both U.S. and Japanese original tweets concerned personal matters (e.g., "I really enjoyed my day today", "今から南アメリカに行くよー✨," translated to "I'm going to South America now ✨"; US: 58.31%, Japanese: 89.82%), and the second most popular category was entertainment (e.g., "Sooooo did Kim Kardashian post today?" "楽しみな映画増えたな," translated to "I'm looking forward to more movies"; US: 25.92%, Japanese: 7.24%), suggesting that the observed results were not due to different topics. To further ensure that observed cultural differences in prevalence were not due to topic content, we reanalyzed only "personal" or "entertainment" tweets, which together comprised over 80% of all tweets for U.S. and Japanese users, and observed the same cultural differences in affective content reported above (see online supplemental materials, Section S3C).

Although our focus was on original posts, we did run similar analyses on retweets, which can be found in online supplemental materials, Sections S4C and S4D.

Discussion

Because most social media research has focused on Western samples, it is unclear whether currently-documented patterns of behavior on social media generalize across the globe. The present research included for the first time both a U.S. and a Japanese sample to test whether cultural affective values might shape what types of affect users produce on social media, as well as what types of affect users are most influenced by in others' posts. This required several improvements upon previous work, including distinguishing between low and high arousal positive and negative states, assessing both affect prevalence and contagion in the same study, controlling for baseline differences in exposure when assessing contagion, and developing a tool for analyzing Japanese sentiment in short text. These innovations reveal that in both the United States and Japan, users tend to produce affective content that *supports* their cultural values but are most influenced by affective content from others that *violates* their cultural values. Because the United States and Japan differ in their affective values, this resulted in cultural differences in the types of affect that users produced the most, as well as differences in the types of affect that users were most influenced by on social media. Whereas U.S. users produced more positive than negative content, Japanese users produced more low arousal than high arousal content, and whereas U.S. users were most influenced by changes in high arousal negative content in others' posts, Japanese users were most influenced by changes in high arousal positive content in others' posts. This pattern of findings held after controlling for potential differences in baseline exposure to different types of affective content as well as topic, and therefore, could not be attributed to these potential confounds.

Values-Violation Account of Virality

The current findings cannot be explained by a more general threat-related account, which would imply that both U.S. and Japanese users should be most influenced by changes in others' high arousal negative content. Instead, our findings support a more culturally specific values-violation account of virality, in which users are most influenced by changes in affect that violate their specific cultural values. Anger, hate, and other high arousal negative states violate the U.S. valuation of positivity, whereas excitement and other high arousal positive states violate the Japanese valuation of low arousal states. Although these findings build on previous work indicating that U.S. users are more likely to share outrage posted by ingroup members (Brady et al., 2017), they further suggest that in countries with different affective values (like Japan), users are instead influenced by different affective states.

We theorize that people may be most influenced by affective content that violates values because violations "hijack" attention (Erber & Fiske, 1984; Mu et al., 2015). Once people attend to values-violating affective content, they may automatically mimic and adjust their emotions to fit that affect, as suggested by emotion contagion theories (Hatfield et al., 1993), or they may actually experience the affective states they are exposed to, as suggested by incidental and anticipatory affect theories (Gummerum et al., 2016; Knutson & Greer, 2008; Loewenstein & Lerner, 2003; Van Dillen et al., 2012). Users then may be more likely to post content that matches this affective content. In addition to attracting attention, increased salience may lead to other attributions about the sender (e.g., increased emotion or veracity) which might also promote transmission. This and other possible processes would be interesting to pursue in future research.

Importantly, the more contagious an affect type was, the less prevalent it was overall in users' posts. This counterintuitive association suggests that multiple mechanisms might influence what people post, and that opposing mechanisms may drive prevalence and contagion. Posts that people produce on their own may be most influenced by their cultural affective values, whereas posts that are a result of exposure may be more due to changes in how people actually feel as a function of exposure. We argue that because cultural values shape what people produce, people are more psychologically sensitive to content that violates those values. An alternative explanation is that users are more sensitive to changes in affect that violate cultural values because they are structurally (vs. psychologically) more novel. Yet another account might posit that users are more sensitive to any type of affect (not just values-violating affect) that is novel or infrequent.

In the present study, these alternative explanations could apply to U. S. users because the prevalence of affect that U.S. users produced and that they were exposed to (i.e., that their follows posted) were similar. Therefore, consistent with the finding that U.S. users were more influenced by affect that was less prevalent among their own tweets, U.S. users were also more influenced by affect that was less prevalent among their follows' tweets. This was not the case for Japanese users, however; the prevalence of affect that Japanese users produced differed from the prevalence of affect that they were exposed to. Thus, although Japanese users were generally more influenced by affect that was less prevalent among their own tweets, they were not more influenced by affect that was less prevalent among their follows' tweets, supporting a values violation account over novelty-based accounts (see online supplementary materials Section S3D). Future research is clearly needed to test these potential mechanisms more directly than was possible in the present study.

Limitations and Future Directions

The findings and limitations of this study generate many new directions for future research. First, potentially interesting information about users (e.g., age, socioeconomic status) could not be accessed via the Twitter API for privacy reasons. Moreover, we could not directly measure users' ideal affect. Smaller-scale studies might recruit specific subsamples to address potential influences of user characteristics, and whether there is a direct link between user's affective values and affective experience with subsequent social media behavior. Second, based on theoretical predictions about cultural differences, this research focused on affective content that varied in terms of valence and arousal dimensions, but other more specific feelings might be of interest in future investigations (e.g., socially engaging vs. disengaging emotions; Kitayama et al., 2006).

Third, like prior research (e.g., Coviello et al., 2014; Ferrara & Yang, 2015; Kramer et al., 2014), we limited our analyses to the text of tweets (including emojis). Many tweets also contain pictures and videos, however, which can even more potently convey affect. To our knowledge, no tools exist to examine the affective content of pictures and videos in tweets and other forms of social media at this scale, but once developed, these tools would allow us to examine whether the current findings generalize to the affective content of pictures and videos. Deconstruction of the content of "original posts" also merits further exploration (e.g., is original content produced in the context of retweeted content subject to the same cultural influences as those that are not?). Future research might build on current findings to address these finer-grained questions.

Fourth, while we went to great lengths to collect representative samples of users, Twitter users themselves are not representative of the general population (Mislove et al., 2011). Despite this, our findings suggest that the sampled individuals were influenced by their culture's affective values. Furthermore, since this research focused on posts, it did not include passive consumers of social media (i.e., users who browse social media but do not post). Similarly, this research did not examine situations in which active consumers of social media decide not to post, and cultural affective values might play a role in abstention as well as production. For instance, compared to the United States, being exposed to high arousal negative content might prevent users from posting any content more in Japan as a way of suppressing or moderating their emotions (Miyamoto et al., 2014; Murata et al., 2013). Thus, future studies might target other types of social media users. We also focused on the United States and Japan based on theoretical predictions and decades of empirical research demonstrating clear differences in the affective values endorsed by members of these cultures. Future studies are of course needed to determine whether these findings generalize to other cultures with different affective values.

Fifth, like prior research (e.g., Coviello et al., 2014; Ferrara & Yang, 2015), we could not determine which posts each user had actually read, and so estimated exposure by aggregating previous posts of users' follows immediately prior to users' original posts. This practice is consistent with recommendations that observational studies conservatively estimate exposure by using 100% of a followed user's content to prevent sampling issues (Morstatter et al., 2013). To control for individual differences in user characteristics as well as Twitter personalization algorithms, our contagion model focused on changes in exposure within users, controlling for users' baseline exposure. However, future work that experimentally manipulates exposure to specific affective content is clearly needed to test causal influence directly.

Similarly, in seeking to capture a typical user's exposure to affective content, we focused on users rather than specific tweets. Future work might complementarily explore a tweet-based "emotion cascade approach" (Goldenberg & Gross, 2020) that facilitates tracking the spread of specific tweets with varying affective content in different cultures. As in prior research on emotional contagion, we focused on how exposure to an affect type predicts likelihood of posting the same or "congruous" affect type; however, there was some evidence of contagion among different or "incongruous" affect pairs. For example, an increase in exposure to LAN affect increased the likelihood of US users posting HAN affect, though to a lesser extent than did an increase in exposure to HAN affect. Exposure to lower levels of values violating affect may increase the likelihood of posting more intense levels of values violating affect. Future research is needed to assess the robustness of these incongruous effects.

Finally, we modeled Japanese SentiStrength after English SentiStrength to code the affective quality of Japanese posts. Although Japanese SentiStrength demonstrated comparable sensitivity for negativity, it showed slightly different sensitivity for positivity compared to English SentiStrength. Because the confusion matrices of Japanese SentiStrength were highly correlated with those of Japanese human raters, however, this may reflect differences between Japanese and English linguistic expression or detection of positive emotion in short text. Thus, cultural differences in the prevalence of high arousal positive content in original posts may result in part from cultural differences in the categorization of low and high positive arousal, which may or may not be related to affective values. Clearly, future research will need to disentangle these possibilities. Furthermore, like other natural language processing programs, SentiStrength is limited in its ability to code semantic meaning (Cambria et al., 2016). Even though SentiStrength includes built-in structural language rules such as negation (e.g., "not" happy) that capture some semantic meaning, future researchers may further develop Japanese SentiStrength to capture the more nuanced meanings and connotations of affective content in text.

Implications for Understanding Culture and Affect on Social Media

This research contributes to the literature on emotion and affect on social media in several ways. First, while the findings replicate previous patterns for U.S. users (Bazarova et al., 2012; Crockett, 2017; Lin & Utz, 2015; Reinecke & Trepte, 2014), they further demonstrate that these patterns do not necessarily generalize to users with different cultural affective values. Yet, these different patterns are still generally interpretable through the lens of supporting and violating cultural ideals. Second, this research suggests a cultural mechanism to explain why people produce specific types of affect on social media, as well as why different types of affect are more viral. Third, the work more generally suggests that how culture influences what people originally produce on social media may differ from how culture influences people's sensitivity to others' content on social media. Fourth, these findings demonstrate the utility and importance of distinguishing low from high arousal positive and negative affective states and treating them as independent, in order to facilitate comparisons among the different affect types. Finally, the findings illustrate the importance of measuring change within users and controlling for differences in baseline exposure to ensure that observed patterns are not due to user similarity or other shared characteristics.

These findings also have broader implications for theories that focus on the intersection of emotion and culture. On the one hand, consistent with affect valuation theory, the overall prevalence of affective content on social media reflects broader cultural affective values, similar to other forms of media (e.g., children's storybooks, magazine advertisements, and leaders' website photos; Tsai, 2007; Tsai et al., 2016; Tsai, Louie, et al., 2007). Thus, although people use social media for many different purposes, these findings demonstrate that social media can provide a clear channel for people to express cultural affective values, even though social media is more dynamic and less deliberately constructed than more traditional forms of media. On the other hand, these findings conversely suggest that the types of affective content that people are most sensitive to and influenced by on social media are those that violate their cultural values. This finding is not consistent with affect valuation theory, and instead suggests that additional mechanistic accounts are needed to understand how viral affective content might "hijack" cultural affective values. Thus, the present research both extends and illustrates a boundary of affect valuation theory.

Practical Implications

As social media becomes a primary channel of communication, broader awareness that people's online behavior reflects their cultural affective values might help reduce common misunderstandings. For instance, affect valuation may be mistaken for affective experience. In the absence of understanding that U.S. posts reflect valuation of positive affect, Japanese might mistakenly underestimate the degree to which Americans feel negative emotions. Conversely, in the absence of understanding that Japanese posts reflect valuation of low arousal affect, Americans might underestimate the degree to which Japanese feel high arousal states.

Even more urgently, these findings might also suggest new ways of combating potentially harmful psychological consequences of social media use. For example, in the United States, social media has been cited as one cause of decreased well-being among young users, in part because viewing peers' posts can make users feel like they are "missing out," or not doing as well as others (Vogel et al., 2014). Such feelings might be mitigated if younger consumers understood that their peers may be producing posts that more closely reflect ideal rather than actual feelings (e.g., users posting to show excitement even when they do not feel excitement). Further, these findings may help combat the harmful effects of social media on society. In the United States, scholars and policymakers alike have raised concerns about the increase in high arousal negative content (anger, hate, moral outrage) on social media, especially in the context of subsequent political polarization, dehumanization of outgroup members, and spread of misinformation (Brady et al., 2017; Crockett, 2017; Vosoughi et al., 2018; Williams, 2018). Our findings, however, suggest that these societal costs could be mitigated if tools were developed to reduce users' exposure to countercultural affect.

References

Barnes, J., Klinger, R., & Schulte im Walde, S. (2017). Assessing state-of-the-art sentiment models on state-of-the-art sentiment datasets. *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media Analysis*.

Barsade, S. G. (2002). The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly*, *47*(4), 644–675. https://doi.org/10.2307/3094912

Bazarova, N. N., Taft, J. G., Choi, Y. H., & Cosley, D. (2012). Managing impressions and relationships on Facebook: Self-presentational and relational concerns revealed through the analysis of language style. *Journal of Language and Social Psychology*, *32*(2), 121–141. https://doi.org/10.1177/0261927X12456384

Bencharit, L. Z., Ho, Y. W., Fung, H. H., Yeung, D. Y., Stephens, N. M., Romero-Canyas, R., & Tsai, J. L. (2019). Should job applicants be excited or calm? The role of culture and ideal affect in employment settings. *Emotion*, *19*(3), 377–401. https://doi.org/10.1037/emo0000444

Boiger, M., Deyne, S. D., & Mesquita, B. (2013). Emotions in "the world": Cultural practices, products, and meanings of anger and shame in two individualist cultures. *Frontiers in Psychology*, *4*, 867. https://doi.org/10.3389/fpsyg.2013.00867

Brady, W. J., & Crockett, M. J. (2019). How effective is online outrage? *Trends in Cognitive Sciences*, *23*(2), 79–80. https://doi.org/10.1016/j.tics.2018.11.004

Brady, W. J., Crockett, M. J., & Van Bavel, J. J. (2020). The MAD Model of Moral Contagion: The role of motivation, attention, and design in the spread of moralized content online. *Perspectives on Psychological Science*, *15*(4), 978–1010. https://doi.org/10.1177/1745691620917336

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(28), 7313–7318. https://doi.org/10.1073/pnas.1618923114

Cambria, E., Poria, S., Bajpai, R., & Schuller, B. (2016, December). SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives. *Proceedings of COLING 2016, the 26th international conference on computational linguistics: Technical papers* (pp. 2666–2677).

Chmiel, A., Sienkiewicz, J., Thelwall, M., Paltoglou, G., Buckley, K., Kappas, A., & Hołyst, J. A. (2011). Collective emotions online and their influence on community life. *PLoS ONE*, *6*(7), e22207. https://doi.org/10.1371/journal.pone.0022207

Clark, A. E., & Kashima, Y. (2007). Stereotypes help people connect with others in the community: A situated functional analysis of the stereotype consistency bias in communication. *Journal of Personality and Social Psychology*, *93*(6), 1028–1039. https://doi.org/10.1037/0022-3514.93.6.1028

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Academic Press.

Coviello, L., Sohn, Y., Kramer, A. D. I., Marlow, C., Franceschetti, M., Christakis, N. A., & Fowler, J. H. (2014). Detecting emotional contagion in massive social networks. *PLoS ONE*, *9*(3), e90315. https://doi.org/10.1371/journal.pone.0090315

Crockett, M. J. (2017). Moral outrage in the digital age. *Nature Human Behaviour*, *1*(11), 769–771. https://doi.org/10.1038/s41562-017-0213-3

Curhan, K. B., Sims, T., Markus, H. R., Kitayama, S., Karasawa, M., Kawakami, N., Love, G. D., Coe, C. L., Miyamoto, Y., & Ryff, C. D. (2014). Just how bad negative affect is for your health depends on culture. *Psychological Science*, *25*(12), 2277–2280. https://doi.org/10.1177/0956797614543802

Erber, R., & Fiske, S. T. (1984). Outcome dependency and attention to inconsistent information. *Journal of Personality and Social Psychology*, *47*(4), 709–726. https://doi.org/10.1037/0022-3514.47.4.709

Feldman-Barrett, L., & Russell, J. A. (1999). The structure of current affect: Controversies and emerging consensus. *Psychological Science*, *8*, 10–14.

Ferrara, E., & Yang, Z. (2015). Measuring emotional contagion in social media. *PLoS ONE*, *10*(11), e0142390. https://doi.org/10.1371/journal.pone.0142390

Goldenberg, A., & Gross, J. J. (2020). Digital emotion contagion. *Trends in Cognitive Sciences*, *24*(4), 316–328. https://doi.org/10.1016/j.tics.2020.01.009

Grossmann, I., Huynh, A. C., & Ellsworth, P. C. (2016). Emotional complexity: Clarifying definitions and cultural correlates. *Journal of Personality and Social Psychology*, *111*(6), 895–916. https://doi.org/10.1037/pspp0000084

Gummerum, M., Van Dillen, L. F., Van Dijk, E., & López-Pérez, B. (2016). Costly third-party interventions: The role of incidental anger and attention focus in punishment of the perpetrator and compensation of the victim. *Journal of Experimental Social Psychology*, *65*, 94–104. https://doi.org/10.1016/j.jesp.2016.04.004

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1993). Emotional contagion. *Current Directions in Psychological Science*, *2*(3), 96–99. https://doi.org/10.1111/1467-8721.ep10770953

Heine, S. J., Lehman, D. R., Markus, H. R., & Kitayama, S. (1999). Is there a universal need for positive self-regard? *Psychological Review*, *106*(4), 766–794. https://doi.org/10.1037/0033-295X.106.4.766

Kelly, J. R., Iannone, N. E., & McCarty, M. K. (2016). Emotional contagion of anger is automatic: An evolutionary explanation. *British Journal of Social Psychology*, *55*(1), 182–191. https://doi.org/10.1111/bjso.12134

Kim, H., & Markus, H. R. (1999). Deviance or uniqueness, harmony or conformity? A cultural analysis. *Journal of Personality and Social Psychology*, *77*(4), 785–800. https://doi.org/10.1037/0022-3514.77.4.785

Kitayama, S., & Markus, H. R. (2000). The pursuit of happiness and the realization of sympathy: Cultural patterns of self, social relations, and well-being. In E. Diener & E. Suh (Eds.), *Subjective well-being across cultures* (pp. 113–161). MIT Press.

Kitayama, S., Mesquita, B., & Karasawa, M. (2006). Cultural affordances and emotional experience: Socially engaging and disengaging emotions in Japan and the United States. *Journal of Personality and Social Psychology*, *91*(5), 890–903. https://doi.org/10.1037/0022-3514.91.5.890

Knutson, B., & Greer, S. M. (2008). Anticipatory affect: neural correlates and consequences for choice. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1511), 3771–3786.

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(24), 8788–8790. https://doi.org/10.1073/pnas.1320040111

Kuppens, P., Ceulemans, E., Timmerman, M. E., Diener, E., & Kim-Prieto, C. H. U. (2006). Universal intracultural and intercultural dimensions of the recalled frequency of emotional experience. *Journal of Cross-Cultural Psychology*, *37*, 491–515.

Kuppens, P., Tuerlinckx, F., Russell, J. A., & Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological Bulletin*, *139*(4), 917–940. https://doi.org/10.1037/a0030811

Lin, R., & Utz, S. (2015). The emotional responses of browsing Facebook: Happiness, envy, and the role of tie strength. *Computers in Human Behavior*, *52*, 29–38. https://doi.org/10.1016/j.chb.2015.04.064

Lin, R., & Utz, S. (2017). Self-disclosure on SNS: Do disclosure intimacy and narrativity influence interpersonal closeness and social attraction? *Computers in Human Behavior*, *70*, 426–436. https://doi.org/10.1016/j.chb.2017.01.012

Loewenstein, G., & Lerner, J. (2003). The role of affect in decision making. In R. Davidson, K. Scherer, & Goldsmith, H. H. (Eds.). *Handbook of affective sciences* (pp, 619–642). Oxford University Press.

Macskassy, S. A., & Michelson, M. (2011). Why do people retweet? Anti-homophily wins the day! In L. Adamic, R. Baeza-Yates, & S. Counts (Eds.), *Proceedings of the Fifth International AAAI Conference on Web and Social Media* (pp. 209–216). Retrieved from https://ojs.aaai.org/index.php/ICWSM/article/view/14110

Markus, H. R., & Kitayama, S. (2010). Cultures and selves: A cycle of mutual constitution. *Perspectives on Psychological Science*, *5*(4), 420–430. https://doi.org/10.1177/1745691610375557

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*, 425–444. https://doi.org/10.1146/annurev.soc.27.1.415

Mislove, A., Lehmann, S., Ahn, Y. Y., Onnela, J. P., & Rosenquist, J. (2011, July). Understanding the demographics of Twitter users. In L. Adamic, R. Baeza-Yates, & S. Counts (Eds.), *Proceedings of the International AAAI Conference on Web and Social Media* (pp. 554–557). Retrieved from https://ojs.aaai.org/index.php/ICWSM/article/view/14168

Miyamoto, Y., & Ma, X. (2011). Dampening or savoring positive emotions: A dialectical cultural script guides emotion regulation. *Emotion*, *11*(6), 1346–1357. https://doi.org/10.1037/a0025135

Miyamoto, Y., Ma, X., & Petermann, A. G. (2014). Cultural differences in hedonic emotion regulation after a negative event. *Emotion*, *14*(4), 804–815. https://doi.org/10.1037/a0036257

Miyamoto, Y., Ma, X., & Wilken, B. (2017). Cultural variation in pro-positive versus balanced systems of emotions. *Current Opinion in Behavioral Sciences*, *15*, 27–32. https://doi.org/10.1016/j.cobeha.2017.05.014

Miyamoto, Y., Uchida, Y., & Ellsworth, P. C. (2010). Culture and mixed emotions: Co-occurrence of positive and negative emotions in Japan and the United States. *Emotion*, *10*(3), 404–415. https://doi.org/10.1037/a0018430

Morstatter, F., Pfeffer, J., Liu, H., & Carley, K. (2013). Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's firehose. In E. Kiciman, N. Ellison, B. Hogan, P. Resnick, & I. Soboroff. (Eds.). *Proceedings of the International AAAI Conference on Web and Social Media* (pp. 400–408). Retrieved from https://ojs.aaai.org/index.php/ICWSM/article/view/14401

Mu, Y., Kitayama, S., Han, S., & Gelfand, M. J. (2015). How culture gets embrained: Cultural differences in event-related potentials of social norm violations. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(50), 15348–15353. https://doi.org/10.1073/pnas.1509839112

Murata, A., Moser, J. S., & Kitayama, S. (2013). Culture shapes electrocortical responses during emotion suppression. *Social Cognitive and Affective Neuroscience*, *8*(5), 595–601. https://doi.org/10.1093/scan/nss036

Oh, S., & Syn, S. Y. (2015). Motivations for sharing information and social support in social media: A comparative analysis of Facebook, Twitter, Delicious, YouTube, and Flickr. *Journal of the Association for Information Science and Technology*, *66*(10), 2045–2060. https://doi.org/10.1002/asi.23320

Park, B., Blevins, E., Knutson, B., & Tsai, J. L. (2017). Neurocultural evidence that ideal affect match promotes giving. *Social Cognitive and Affective Neuroscience*, *12*(7), 1083–1096. https://doi.org/10.1093/scan/nsx047

Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). *Linguistic inquiry and word count: LIWC2015*. Pennebaker Conglomerates.

Reinecke, L., & Trepte, S. (2014). Authenticity and well-being on social network sites: A two-wave longitudinal study on the effects of online authenticity and the positivity bias in SNS communication. *Computers in Human Behavior*, *30*, 95–102. https://doi.org/10.1016/j.chb.2013.07.030

Ruby, M. B., Falk, C. F., Heine, S. J., Villa, C., & Silberstein, O. (2012). Not all collectivisms are equal: Opposing preferences for ideal affect between East Asians and Mexicans. *Emotion*, *12*(6), 1206–1209. https://doi.org/10.1037/a0029118

Simpson, A., & Kashima, Y. (2013). How can a stereotype inconsistency bias be encouraged in communication? *Asian Journal of Social Psychology*, *16*(1), 71–78. https://doi.org/10.1111/ajsp.12010

Sims, T., Tsai, J. L., Jiang, D., Wang, Y., Fung, H. H., & Zhang, X. (2015). Wanting to maximize the positive and minimize the negative: Implications for mixed affective experience in American and Chinese contexts. *Journal of Personality and Social Psychology*, *109*(2), 292–315. https://doi.org/10.1037/a0039276

Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C., Ng, A., & Potts, C. (2013). Recursive deep models for semnatic compositionality over a sentiment treebank. In D. Yarowsky, T. Baldwin, A. Korhonen, K. Livescu, & S. Bethard (Eds.), *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (pp. 1631–1642). Association for Computational Linguistics. https://aclanthology.org/D13-1170.pdf

Thelwall, M. (2017). Heart and soul: Sentiment strength detection in the social web with SentiStrength (summary book chapter). In J. Holyst (Ed.), *Cyberemotions: Collective emotions in cyberspace* (pp. 119–134). Springer. https://doi.org/10.1007/978-3-319-43639-5_7

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, *61*(12), 2544–2558. https://doi.org/10.1002/asi.21416

Tsai, J. L. (2007). Ideal affect: Cultural causes and behavioral consequences. *Perspectives on Psychological Science*, *2*(3), 242–259. https://doi.org/10.1111/j.1745-6916.2007.00043.x

Tsai, J. L. (2017). Ideal affect in daily life: Implications for affective experience, health, and social behavior. *Current Opinion in Psychology*, *17*, 118–128. https://doi.org/10.1016/j.copsyc.2017.07.004

Tsai, J. L., Ang, J. Y. Z., Blevins, E., Goernandt, J., Fung, H. H., Jiang, D., Elliott, J., Kölzer, A., Uchida, Y., Lee, Y.-C., Lin, Y., Zhang, X., Govindama, Y., & Haddouk, L. (2016). Leaders' smiles reflect cultural differences in ideal affect. *Emotion*, *16*(2), 183–195. https://doi.org/10.1037/emo0000133

Tsai, J. L., Blevins, E., Bencharit, L. Z., Chim, L., Fung, H. H., & Yeung, D. Y. (2019). Cultural variation in social judgments of smiles: The role of ideal affect. *Journal of Personality and Social Psychology*, *116*(6), 966–988. https://doi.org/10.1037/pspp0000192

Tsai, J. L., & Clobert, M. (2019). Cultural influences on emotions: Established patterns and emerging trends. To appear. In S. Kitayama & D. Cohen (Eds.), *Handbook of cultural psychology* (2nd ed.). Guilford Press.

Tsai, J. L., Knutson, B., & Fung, H. H. (2006). Cultural variation in affect valuation. *Journal of Personality and Social Psychology*, *90*, 288–307.

Tsai, J. L., Louie, J. Y., Chen, E. E., & Uchida, Y. (2007). Learning what feelings to desire: Socialization of ideal affect through children's storybooks. *Personality and Social Psychology Bulletin*, *33*(1), 17–30. https://doi.org/10.1177/0146167206292749

Tsai, J. L., Miao, F. F., Seppala, E., Fung, H. H., & Yeung, D. Y. (2007). Influence and adjustment goals: Sources of cultural differences in ideal affect. *Journal of Personality and Social Psychology*, *92*(6), 1102–1117. https://doi.org/10.1037/0022-3514.92.6.1102

Tsukawaki, R., Fukada, H., & Higuchi, M. (2011). Process effects of expression of humor on anxiety and depression. *Japanese Journal of Experimental Social Psychology*, *51*(1), 43–51. https://doi.org/10.2130/jjesp.51.43

Van Dillen, L. F., van der Wal, R. C., & van den Bos, K. (2012). On the role of attention and emotion in morality: Attentional control modulates unrelated disgust in moral judgments. *Personality and Social Psychology Bulletin*, *38*(9), 1222–1231. https://doi.org/10.1177/0146167212448485

Vogel, E. A., Rose, J. P., Roberts, L. R., & Eckles, K. (2014). Social comparison, social media, and self-esteem. *Psychology of Popular Media Culture*, *3*(4), 206–222. https://doi.org/10.1037/ppm0000047

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, *98*(2), 219–235. https://doi.org/10.1037/0033-2909.98.2.219

Williams, Z. (2018). *Why are we living in an age of anger—is it because of the 50-year rage cycle?* https://www.theguardian.com/science/2018/may/16/living-in-an-age-of-anger-50-year-rage-cycle

Yik, M., & Russell, J. A. (2003). Chinese affect circumplex: I. Structure of recalled momentary affect. *Asian Journal of Social Psychology*, *6*(3), 185–200. https://doi.org/10.1046/j.1467-839X.2003.00120.x

Yoshida, T., Kojo, K., & Kaku, H. (1982). A study of the development of self-presentation in children. *Japanese Journal of Educational Psychology*, *30*, 2030–2037. https://doi.org/10.5926/jjep1953.30.2_120

Received September 4, 2020
Revision received April 28, 2021
Accepted May 5, 2021 ■