Received: 10 January 2022 | Revised: 22 September 2022 | Accepted: 29 November 2022

DOI: 10.1002/col.22837

RESEARCH ARTICLE

COLOR RESEARCH AND APPLICATION WILEY

Sentiment analysis based on frequency of color names on social media

Boshuo Guo | **Stephen Westland** | **Peihua Lai**

School of Design, The University of Leeds, Leeds, LS2 9JT, UK

Correspondence
Boshuo Guo, The University of Leeds, Leeds, UK.
Email: b.guo1@leeds.ac.uk

Abstract

This study explores the temporal changes in sentiment associated with eight color names over an 18-month period at four observation points. We focus on the valence aspect of sentiment. We collected four datasets, each separated by 6 months, and each containing 18 000 mentions of each of the eight color names in English from Twitter users around the world. We calculated the weighted average sentiment score of each instance when a color is mentioned. We find that purple and pink are the most positive in average sentiment score in all observation points, whereas brown, red, and orange are ranked as the lowest in average sentiment score. In terms of relative rank in sentiment value associated with the color names, we find the three consecutive datasets of July 2020, January 2021 and July 2021 are more consistent with one another, while the January 2022 dataset is more different from the earlier three datasets. This finding indicates that the temporal consistency in color-associated sentiment might maintain within 1 year, while evolve and show more difference in a longer timeline. This study is useful to marketing professionals by revealing that color names are associated with sentiment and that these associations can be monitored using social media data regularly. We suggest that marketers can use our method to analyse the color-associated sentiment of color names regularly, maybe on an annual basis, in order to choose color names wisely.

KEYWORDS

color, marketing, psychology

1 | INTRODUCTION

Color plays an important role in marketing communication. Consumers associate colors with specific words[1] and feelings,[2] which may influence their attitudes towards advertisements, brands and products that use those colors.[2] Research shows that color can make advertisements more attractive,[3] promote favorable attitudes[4,5] and influence consumers' information processing[6] and their responses to advertising.[1,7] Understanding the sentiment associated with colors may enable marketers to generate positive online word of mouth references, improve persuasion[8] and communicate effectively with customers.[7] The fact that colors convey meanings is undoubtedly a key aspect of their power in advertising and marketing. Color meanings have been widely studied although there is still uncertainty about the extent to which color meanings are generic and the extent to which they depend upon context.[9] Kress and Van Leeuwen[10] have argued that social groups that share common purposes around color are often relatively small and specialized compared to groups who share speech or visual communication. Indeed, it has even been suggested that color per se does not elicit response, but that the particular meaning or significance of a color is context-bound and varies from one person or situation to another.[11] Despite this theoretical idea, empirical data tend to suggest otherwise and provide some evidence for a degree of universality of color meanings[9,12] even if it is clear that color meanings in practice often depend upon the context in which the color is used. Although a great many studies find that colors affect consumers[13–15] it is nonetheless difficult to find simple and clear guidelines as to exactly *how* color affects consumers. This may be because the meanings that colors have often vary depending upon the context. For example, the meaning of the color red when used in a stop sign is quite different from the meaning of the same color when used on the packaging of a foodstuff such as fruit yoghurt.

Past studies on the effects of color in marketing communications have mostly focused on the meanings elicited by the visual components of color: namely, hue, lightness, and saturation.[1,2,5,7,16] Far fewer studies, if any, have considered the effects of color in marketing in terms of literal color names. There has, of course, been a great deal of research into the relationship between color names and color perception. Many of these studies have been based on the idea[17] of their being 11 basic color names (black, white, red, green, yellow, blue, pink, gray, brown, orange and purple) which elicit color sensations that are shared across all cultures. For example, Lin et al.[18] developed a color-naming model to categorize all color coordinates in CIELAB color space into 11 basic color names. The concept of basic color names has been used to argue for color universality; that is, that despite our different languages and cultures our perception of color is essentially universal. This contrasts with linguistic relativity (sometimes expressed as the Sapir–Whorf hypothesis) that the language that we speak affects our worldview and, in the context of color, our color perception. Today, there is still some controversy about the universalist approach and a number of studies have somewhat contracted the work of Berlin and Kay.[19] Recently, some studies have reported interesting online experiments to enable studies to be completed with large numbers of participants.[20,21]

However, this work is not concerned with the relationship between color names and color perception but rather focusses on the meanings and emotions that may be elicited by the color names themselves. Since color names are frequently used in brand names (e.g., Red Bull) and marketing campaigns, marketers need to understand the sentiment associated to literal names of color in order to make informed choices when selecting color names in marketing communication.

In this study, we aim at finding out the sentiment associated with eight color names and how the sentiment varies over time 2020–2022. We use data collected from social media to measure the sentiment. Sentiment analysis refers to identifying as well as classifying the sentiments that are expressed in a text source. A large number of studies have been carried out to analyze text from social media such as Twitter because social media platforms provide huge amounts of text and access to text from a large number of people (consumers). Sentiment analysis over Twitter allows a fast and effective way to monitor consumers' feelings towards a brand, product, or business. For example, one study analyzed tens of thousands of tweets to determine sentiments for a number of political issues in the USA such as Health Care Reform and the Obama-McCain debate.[22] Typically, tweets are categorized as being either positive or negative (sometimes there is a neutral category) and reviews of methods to achieve this have been published.[23,24]

In the next sections, we describe our data, develop our measurement of sentiment and discuss our findings, as well as study limitations and directions for future research.

2 | DATA AND METHOD

2.1 | Data Collection

The theoretical background of this study is word-color association[1] and research in English color lexicon.[17,25] Research in English color lexicon proposed 11 basic color names according to five dimensions: the term is monolexemic; it is used to principally refer to color of things; it is not constrained to describing color of a specific object; it is used by most of English speakers and it can be used to partition color space exhaustively.[17] According to the five dimensions, Berlin and Kay[17] proposed blue, brown, green, orange, pink, purple, red, yellow, black, white and grey as the 11 basic color names. Following the studies in English color lexicon, our research target chromatic terms that are within these 11 basic color names. Among the 11 basic color names,[17] we focus on the eight colors names blue, brown, green, orange, pink, purple, red and yellow, following Dorcus' pioneering research on word-color association.[1] We did not consider black, white or grey, as, according to the construal-level theory in

TABLE 1 Number of unique users who mentioned each color in posts

Time	Blue	Brown	Green	Orange	Pink	Purple	Red	Yellow	Total
07/2020	12 145	14 414	12 998	13 272	12 987	12 273	14 754	14 553	101 702
01/2021	15 489	13 477	15 597	14 239	14 556	13 642	15 379	14 728	112 045
07/2021	15 249	14 872	15 497	15 046	14 787	14 107	15 567	14 120	107 615
01/2022	14 891	15 034	15 312	14 930	14 518	13 489	15 761	14 899	108 629

consumer research, imagery of these three colors promotes high-level construal, while the other colors promotes low-level construal.[6,26] As different construal levels promote different emotions,[27] we exclude these three colors to avoid potential bias. Although we focus on color lexicon instead of imagery, this difference in construal-level might be over-generalised to lexicon and it is prudent to avoid this potential effect.

We collected four datasets in July 2020, January 2021, July 2021, and January 2022. Each dataset contains 18 000 mentions of each of the eight color names in English from Twitter users around the world. We collected the datasets from Twitter, using the Twitter API and the Rtweet[28] package. Each dataset contains 144 000 Twitter posts (18 000 posts per color name $\times$ 8 color names). Every post mentions at least one of the eight color names in the text of the post. In order to avoid repeated content in retweets and consequent bias in calculation, we included only original Twitter posts in our datasets (no retweets are included). We group the posts according to mentions of color.

We collected the four datasets at the four different observation times, in order to assess the coherence and temporal consistency in the sentiment associated with these color names. Table 1 lists the observation time and the numbers of users of every dataset.

The posts mentioning each color were collected from different groups of users. That is, taking the July 2020 and the January 2021 datasets as examples, we collected the twitter posts that mentioned one color from a group of users in July 2020 and the twitter posts that mentioned the same color from a totally different group of users in January 2021. This research design is to remove the confounding effect of individual-level temporal consistency in color-associated sentiment from the overall consistency of the color-associated sentiment that we want to test.

In our datasets, every post is described with 90 variables.[28] These variables include the time when the post was created, the post text, the user's name, hashtags, number of likes and more. However, we simply focus our analysis on the post text in this study given our key interests in the sentiment associated with the literal name of colors. We collected the posts from Twitter users around the world and only include in our datasets the posts in English.

We cleaned the datasets and processed the post text into tidy text[29] using the following steps. First, we removed URLs, and any mentioned user's names and hashtags from the text. This is to avoid any bias in calculating the sentiment value, since, although some URLs and users' names contain words that carry sentiment meanings, quoting these elements does not uncover the intended sentiment of the post author. Stop words are common words that do not carry meaning; for example, "is" and "the." We removed stop words from the text since these words do not carry sentiment meanings. We split the remaining text into single words and calculated the frequency of every word in the 18 000 posts of each color within each dataset.

2.2 | Development of sentiment measurement

We focus on the valence dimension of sentiment. An empirically well-established framework[30] suggests that emotional experiences are best characterized in two dimensions.[31] One dimension is the valence, or pleasure-displeasure dimension, which ranges from highly positive to highly negative.[30,31] In Russell's map[30] that depicts emotional experience, valence describes a person's current feeling. Valence dimension ranges from one extreme negative point (e.g., agony) through a neutral point to a most positive point, such as ecstasy.[30] The other dimension is the arousal dimension, or activating-deactivating dimension.[30] In this study we focus on valence dimension, as our datasets are twitter posts and it is feasible to numerically measure valence from texts using well-developed lexicon dictionaries (e.g., AFINN).

We measure sentiment using the weighted average sentiment score of each instance (or tweet) when a color is mentioned based on the aggregate known sentiment scores of other words in the tweet. We derive sentiment value of each word using the AFINN (Nielsen, 2011) lexicon. AFINN lexicon rates the valence of English word using a score from -5 to $+5$. A positive score indicates positive emotion, while a negative score indicates negative sentiment. A score with a bigger absolute value indicates a stronger emotion. The words in AFINN lexicon were manually scored by the developer of AFINN lexicon (Nielsen, 2011) according to the valence of each word. AFINN lexicon was constructed with words from a range of resources, including twitter postings, The Balanced Affective Word List (Greg)[32] and The Compass DeRose Guide to Emotion Words.[33] Detailed information about the construction and development of AFINN lexicon can be find at (Nielsen, 2011). One strong advantage of using the AFINN lexicon is this lexicon include Internet slangs and acronyms that are widely used in Twitter posts (Nielsen, 2011), which makes AFINN lexicon particularly suitable in analysing the color-associated sentiment in Twitter. Another advantage of AFINN lexicon is that it scores the valence of each word with a numeric value, which enable us to compute the sentiment score of every twitter post. We therefore are able to calculate the overall sentiment score associated with each color.

We identified the 10 most frequently used words and their sentiment scores in the group of posts of each color in each of the datasets. If we consider the color blue as an example, in the twitter posts that mention blue, the five most frequently used positive words are "win, love, lol, beautiful, true"; "love, lol, true, support, matter"; "love, lol, beautiful, top, hope" and "love, lol, beautiful, top, yeah" in the datasets collected in July 2020, January 2021, July 2021 and January 2022 respectively. If a word is used more often in mentions a specific color, it contributes more to the overall sentiment score of the color. Figure A1 displays words that contribute the most sentiment score in the posts of every color.

We develop the measurement of color-associated sentiment as a weighted average sentiment score, as shown in Equation.[23]

$$colour-associated\ sentiment\ Y_i = \frac{\sum_{j=1}^{J}\left(x_j * freq_{ji}\right)}{\sum_{j=1}^{J}\left(freq_{ji}\right)}, \quad (1)$$

Assuming there are J words in all posts mentioning a color i, x_j denotes the sentiment score of the Word j; $freq_{ji}$ denotes the number of times (frequency) for which the Word j was used in all posts mentioning Color i. We calculate the color associated sentiment of color i as Y_i as documented with Equation 1. We use the word frequency as the weight, since a word contributes more to the overall sentiment value if it is used more often in the posts. We multiply every word's sentiment score by the number of times this word was used in all the posts mentioning a certain color. We sum up this value of all the words in the posts mentioning a certain color and divided it by the total word frequency, which follows Nielsen's method of calculating combined sentiment of a tweet (Nielsen, 2011). This approach is beneficial in two ways. First, it is easy to use and adapt. Using any matrix containing word and word frequency, a document-term matrix for example, users can easily use this method to compute the associated sentiment score of a given color. The method could be easily adapted to use with not only text obtained from twitter posts but also a wide range of other texts. The second advantage is that our method considers the word frequency, which enables the color sentiment score Y_i to reflect the relative importance of every word.

3 | RESULTS

3.1 | Temporal consistency in overall color sentiment in datasets of July 2020, January 2021 and July 2021

We present the color associated sentiment values in Figure 1.

We find a temporal consistency within each color as is visually evident in the plots in Figure 1. Purple is the most positively mentioned color in July 2020, January 2021 and July 2021, since purple has the highest overall sentiment score in these datasets. Purple, pink and green are the top three most positive colors, as they are ranked as the first, second and third highest in weighted average sentiment scores in the three datasets.

Blue and yellow are ranked next to purple, pink and green in weighted average sentiment scores. Blue and yellow are positive in the July 2020 and July 2021 datasets while negative in the January 2021 dataset. However, the overall sentiment scores of both colors are small in absolute values, indicating that they are approximately neutral in the three datasets.

Red, brown and orange are negative in both datasets of July 2020 and January 2021. Although red and orange turned positive in July 2021 dataset, these three colors remain ranked at the bottom in weighted average sentiment scores in the three observation points.

The above coherences in the three datasets of July 2020, January 2021 and July 2021 indicate that Twitter users express the most positive sentiment while mentioning purple, pink and green; while they express least positive sentiment while mentioning brown, orange and red in the three observation points.

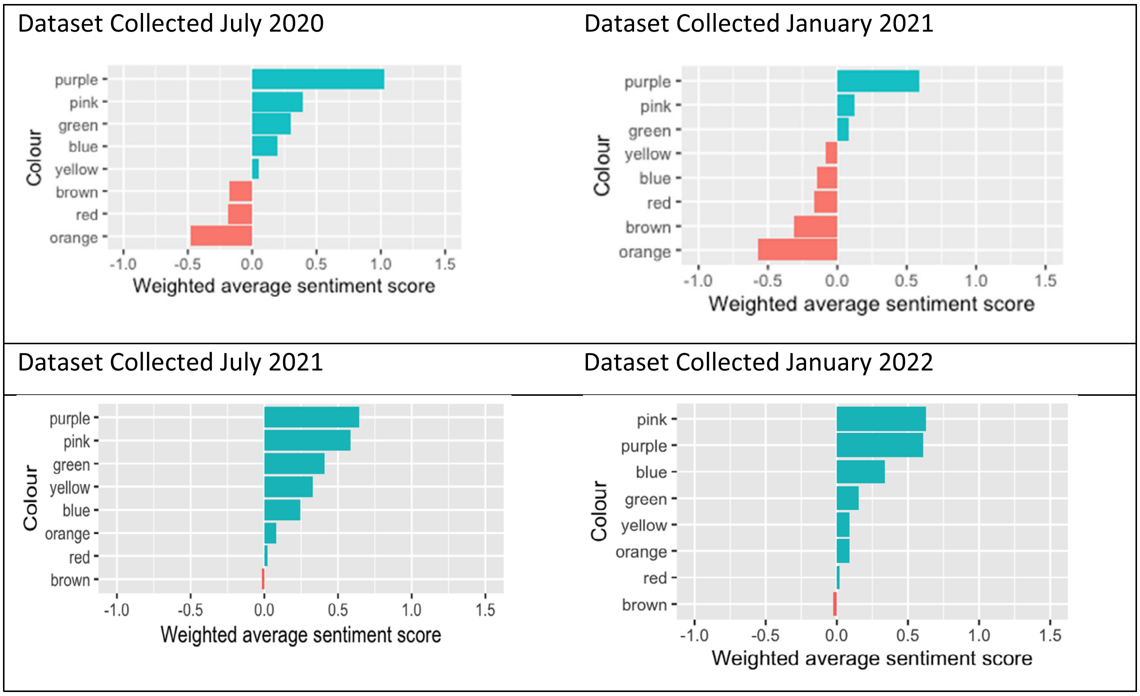

FIGURE 1 Mean sentiment scores for each of the color names

3.2 | Findings in overall color sentiment across all four datasets

Comparing the four datasets, we find that the July 2021 and January 2022 datasets are more positive than the earlier two datasets, as more colors turn to be positive in weighted average sentiment scores in these datasets. It indicates that, compared with the two earlier observation points, Twitter users expressed more positive sentiments when mentioning all eight colors in July 2021 and January 2022 observation points.

We also find that, in terms of the most positive colors, the three consecutive datasets of July 2020, January 2021 and July 2021 are more consistent with each other, as purple, pink and green are always ranked as the first, second and third most positive colors. This temporal consistency does not remain in the dataset of January 2022 where the top three most positive colors are pink, purple and blue.

Besides, in terms of the relative neutral colors, the three consecutive datasets of July 2020, January 2021 and July 2021 are also overall consistent with one another, since blue and yellow are ranked in the middle: they are neither the most positive colors nor ranked in the bottom. The dataset of January 2022 is not so closely consistent where the relatively neutral colors are green and yellow.

In terms of the colors of the lowest weighted average sentiment scores, the four datasets are consistent with one another, as brown, red and orange are ranked at the bottom in all four datasets, although the ranking among the three colors varies.

To summarize, we find the three consecutive datasets of July 2020, January 2021 and July 2021 are more consistent with one another, while the January 2022 dataset is more different from the earlier three datasets. This finding indicates that the temporal consistency in color-associated sentiment might maintain within 1 year (e.g., from July 2020 to July 2021), while evolve and show more difference in a longer timeline, for example, 1 year and a half.

One managerial implication of our finding is regular sentiment analysis. We suggest that marketers, before using a color name to feature a marketing campaign, can use our method to analyse the color-associated sentiment of color names regularly, maybe on an annual basis, in order to guarantee their chosen color is associated with positive sentiment.

Although there is some inconsistency between January 2022 dataset and the earlier datasets, we can still observe one overall consistency through all four datasets. The consistency is that purple and pink are the most positive colors. This finding indicates that, although the ranking among colors might vary over a longer timeline, for example, 1 year and a half, sentiment analysis on a regular basis can enable managers and designers to identify inter-temporal consistency in colors-associated sentiment and monitor the changes.

TABLE 2 Percentage of matched sentiment words between two neighboring observation points

Neighboring Obs. points	Blue	Brown	Green	Orange	Pink	Purple	Red	Yellow
July 2020–January 2021	0.845	0.903	0.869	0.841	0.755	0.775	0.873	0.838
January 2021–July 2021	0.859	0.896	0.871	0.934	0.899	0.866	0.877	0.866
July 2021–January 2022	0.875	0.904	0.902	0.886	0.853	0.863	0.897	0.867
Average	0.860	0.901	0.881	0.887	0.835	0.835	0.882	0.857

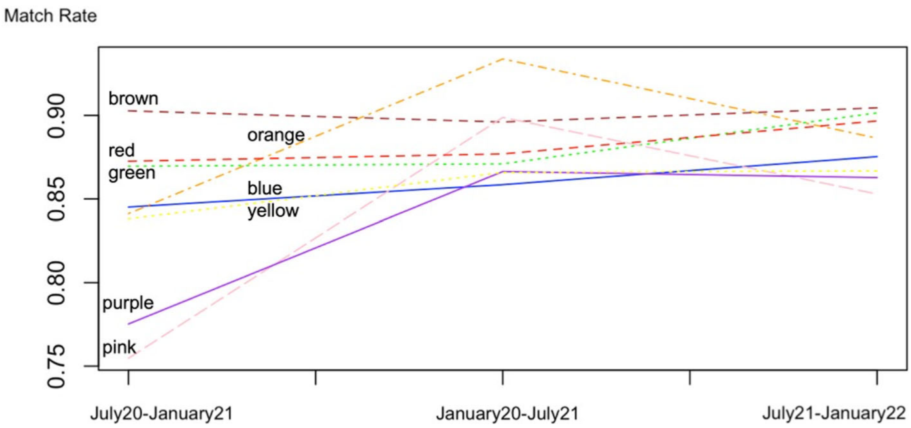

FIGURE 2 Percentage of matched sentiment words between two neighboring observation points

3.3 | Temporal consistency in words

We find that, in all four datasets, purple is the only color where the top 10 frequently used words are all positive, as illustrated in Figure A1. Among the other colors that have both positive words and negative words, pink is the only color that has more positive words than negative in all four datasets. This finding also tells why purple and pink are consistently the most positive color in the four observation points.

In order to better understand the consistency and changes in sentiment words in all the colors, we derive word stem of every word of the Twitter posts in all four datasets. According to matched word stem, we counted the number sentiment words that are identical between every two observation points of every color. We calculate the match rate of the sentiment words between two neighboring observation points (*Obs.t and Obs.[t-1]*) as Equation 2:

$$match\ rate = \frac{number\ of\ sentiment\ words\ matched\ between\ Obs.t\ and\ Obs.(t-1)}{total\ number\ of\ sentiment\ words\ at\ Obs.t}, \tag{2}$$

We define this match rate as the percentage of words in each dataset that can find a match in the previous dataset of a specific color. We summarize the calculated match rates in Table 2:

The Average row in Table 2 shows an overall very high average match rate, as the average match rate of all of the colors are higher than 0.8. The lowest match rate between neighboring observation points is about 0.76. This high average match rate indicates that, when mentioning a color, Twitter users used more than 76% words that are identical with the twitter posts mentioning the same color, collected 6 months ago. This high match rate further supports the feasibility of the managerial implication that reliably monitoring and predicting the color-associate sentiment using Twitter data is possible.

The other finding, the consistency of the colors of the lowest weighted average sentiment scores, can also find further supports from Table 2. Brown, orange and red are found as the colors of the lowest weighted average sentiment scores in all four datasets. These three colors also have the highest average match rates, as shown in Table 2. The average match rates indicate that, in terms of sentiment words used in mentioning each color, these colors are more consistent than the other colors in the four datasets. Twitter users use more identical words in mentions of these three colors, compared with mentions of positive or neutral colors.

Table 2 also provides new insights into the findings about the two most positive colors: purple and pink. To facilitate the interpretation of Table 2, we plot the match rates of the sentiment words between two neighboring observation points of every color in Figure 2.

In Figure 2, the statistics of each color are represented with a line in that color. For example, the solid purple curve represents the match rates of sentiment words between two neighboring observation points in the twitter posts mentioning the color purple, while the dashed pink curve stands for the color pink. Figure 2 depict that, compared with the other colors, the match rate of purple and pink are relatively lower between July 2020 and January 2021 and between July 2021 and January 2022. This relatively lower matching rate indicates that more diverse words were used in Twitter posts mentioning purple and pink than other colors. Although the words are more diverse in purple and pink, the majority of words mentioning the two colors still reflects positive sentiment, since purple and pink are most positive in overall sentiment in all four datasets.

4 | DISCUSSION

Our study uncovers the inter-temporal coherence and changes of sentiment associated with eight chromatic color names in Twitter at four consecutive observation points over a time scope of 1 year and a half. To the best of our knowledge, this study is the first to explore the inter-temporal coherence and changes in sentiment associated to color names in social media using natural language processing methods. Our findings suggest that it is possible to develop tools to automatically monitor and predict the sentiment associated with color names in social media.

Our results identify temporal consistency in overall color sentiment in datasets of July 2020, January 2021 and July 2021. We find that the overall color-associated sentiment in January 2022 dataset is more different from the earlier three datasets. We also find that, when mentioning a color, Twitter users used more than 76% words that are identical with the twitter posts mentioning the same color, collected 6 months ago.

Our findings suggest that monitoring and predicting the color-associate sentiment using Twitter data is feasible. Marketers can analyse the color-associated sentiment regularly, maybe on a six-month basis or an annual basis, to guarantee their chosen color is associated with positive sentiment if they use a color name in marketing campaigns.

As The United States does not only have the biggest number of Twitter users[34] around the world but also contributes the biggest percentage of web traffic[35] to Twitter, we can reasonably assume that the posts of U.S. Twitter users comprise the biggest part of our data. We accordingly suggest that the managerial implication that we suggest is most applicable to U.S. market, specifically, American English-speaker Twitter users.

Our study contributes to marketing communication by suggesting that marketers, before using a color names to feature grammatical marketing communications on social media, can analysis the sentiment associated with this color name on this social media platform to make more cautious managerial decisions on the marketing communication. We provide a procedure that marketers can use easily to compute the overall sentiment value associated with color names.

Our study also contributes to color research by suggesting that English basic color names, when mentioned in social media, can associate with sentiment. The color-associate sentiment is relatively consistent within 1 year and gradually change over a longer time scope. Future studies in associated sentiment of English basic color names can further extend the longitudinal time scope and explore how color-associated sentiment evolves over a longer time scope. In our study, we derive sentiment scores from mentions of colors using data collected from Twitter. Further studies can also test our results using online experiments.

There are several limitations in our study. First, the sentiment scores are only based on the data collected from Twitter. According to the past studies in consumer sentiment analysis, brand sentiment depends on the social media platforms. We assume that color-associated sentiment metric can also depend on the platform. Although we have observed the temporal consistency at one social media platform, further research can be developed on a more comprehensive analysis of data collected from multiple social media platforms.

Second, we focus only on the grammatical mention of color names in text. Since processing text is considered as at a higher construal level in consumer information processing while pictures are at a lower construal level, we assume that the color-associated sentiment metric can be different in pictures. Further studies can compare the color sentiment associated with the color featured photos on social media with the sentiment associated with grammatical mention of color names.

Third, our study is not limited to the color sentiment in any specific country or specific industry. Past studies suggest that the meaning of colors depends on country, culture and industry.[36] Studies by Mylonas and colleagues (Mylonas et al., 2015) have developed successful methods of estimating geo-location information of Twitter posts, if the Twitter posts' geo-code is not hidden. Due to the nature of our datasets, however, the geo-location information is hidden from us and therefore it is not possible for us to recognize the geo-location of the twitter users in this study. As the variation of color-associated sentiment among different countries is important for marketing professionals, future studies could develop a more comprehensive analysis on color sentiment in different geological locations.

Fourth, although we randomly select the twitter posts for this study in order to avoid biases of specific grammatical context, there could be ambiguation in the meaning of colors in our dataset. For example, in a specific grammatical context of food, orange can refer to a fruit instead of a color (Mylonas et al., 2015). Although Part of Speech Analysis (POS) could be used to identify whether a color is used as a noun or an adjective, it is still challenging to avoid this ambiguation since Twitter posts are less than 140 words and are usually in none-standard English (Mylonas et al., 2015). We acknowledge that there is the possibility that our final datasets have irrelevant twitter posts that are not about the colors. However, these possible irrelevant posts, if any, take only a small percentage in the final dataset of 576 000 twitter posts and is unlikely to influence the main results. Future studies could develop more accurate procedure of refining text data and computing color-associated sentiment.

AUTHOR CONTRIBUTIONS

Boshuo Guo: Conceptualization, Methodology, Programming, Data analysis, Visualization, Investigation, Writing, Original draft preparation. **Stephen Westland:** Conceptualization, Investigation, Writing, Reviewing and Editing. **Peihua Lai:** Investigation.

ACKNOWLEDGMENTS

The authors would like to thank Professor Anna Marotta, Professor Larissa Noury and the participants at color and communication/marketing session, AIC 2021, for the insightful comments that have significantly improved the quality of this paper.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request

ORCID

Boshuo Guo https://orcid.org/0000-0001-9940-2998
Stephen Westland https://orcid.org/0000-0003-3480-4755
Peihua Lai https://orcid.org/0000-0002-8095-5928

REFERENCES

[1] Dorcus D. Habitual word associations to colors as a possible factor in advertising. *J Appl Psychol*. 1932;16(3):277-287. doi:10.1037/h0070334

[2] Gorn GJ, Chattopadhyay A, Yi T, Dahl DW. Effects of color as an executional cue in advertising: They're in the shade. *Management Sci*. 1997;43(10):1387-1400.

[3] Bohle, R. H., and Garcia, M. R. (1986). *Reader Reactions to Color in Newspapers* (ED270802). ERIC. https://eric.ed.gov/?id=ED270802

[4] Fernandez KV, Rosen DL. The effectiveness of information and color in yellow pages advertising. *J Advert*. 2000;29(2):61-73.

[5] Meyers-Levy J, Peracchio LA. Understanding the effects of color: how the correspondence between available and required resources affects attitudes. *J Consum Res*. 1995;22(2):121-138.

[6] Lee H, Deng X, Unnava HR, Fujita K. Monochrome forests and colorful trees: the effect of black-and-white versus color imagery on construal level. *J Consum Res*. 2014;41(4):1015-1032.

[7] Sokolik K, Magee RG, Ivory JD. Red-hot and ice-cold web ads: the influence of web ads' warm and cool colors on click-through rates. *J Interact Adv*. 2014;14(1):31-37.

[8] Han D, Duhachek A, Agrawal N. Emotions shape decisions through construal level: the case of guilt and shame. *J Consum Res*. 2014;41(4):1047-1064.

[9] Won S, Westland S. Color meaning in context. *Color Res Appl*. 2017;42(4):450-459.

[10] Kress G, van Leeuwen T. *Reading Images: the Grammar of Visual Design*. Routledge; 2006.

[11] Grieve KW. Traditional beliefs and color perception. *Percept Mot Skills*. 1991;72:1319-1323.

[12] Gao X-P, Xin JH, Sato T, et al. Analysis of cross–-cultural color emotion. *Color Res Appl*. 2007;32(3):223-229.

[13] Spence C, Velasco C. On the multiple effects of packaging color on consumer behaviour and product experience in the 'food and beverage' and 'home and personal care' categories. *Food Quality Preference*. 2018;68:226-237.

[14] Westerman SJ, Sutherland EJ, Gardner PH, et al. Ecommerce interface color and consumer decision making: two routes of influence. *Color Res Appl*. 2012;37(4):292-301.

[15] Yu L, Westland S, Li Z, Xia G. The effect of decision time-length condition on consumer product-color purchase decision. *Color Res Appl*. 2021;46(6):1360-1370.

[16] Lichtlé MC. The effect of an advertisement's color on emotions evoked by attitude towards the ad: the moderating role of the optimal stimulation level. *Int J Ad*. 2007;26(1):37-62.

[17] Berlin B, Kay P. *Basic Color Terms: their Universality and Evolution*. University of California Press; 1969.

[18] Lin H, Luo MR, MacDonald LW, Tarrant AW. A cross-cultural color-naming study. Part III—A color-naming model. *Color Res Appl*. 2001;26(4):270-277.

[19] Jameson KA, Alvarado N. Differences in color naming and color salience in Vietnamese and English. *Color Res Appl*. 2003;28(2):113-138.

[20] Moroney N. Unconstrained web-based color naming experiment, proceedings of SPIE - the international society for. *Opt Eng*. 2003;5008:36-46. doi:10.1117/12.472013

[21] Mylonas D, Griffin LD. Coherence of achromatic, primary and basic classes of color categories. *Vision Res*. 2020;175:14-22.

[22] Saif H, He Y, Alani H. Semantic sentiment analysis of twitter. *International Semantic Web Conference*. Springer; 2012: 508-524.

[23] Antonakaki D, Fragopoulou P, Ioannidis S. A survey of twitter research: data model, graph structure, sentiment analysis and attacks. *Expert Sys Appl*. 2021;164:114006.

[24] Go A, Huang L, Bhayani R. Twitter sentiment analysis. *Entropy*. 2009;17:252.
[25] Lindsey DT, Brown AM. The color lexicon of American English. *J Vis*. 2014;14(2):17.
[26] Liberman N, Trope Y. The psychology of transcending the here and now. *Science*. 2008;322(5905):1201-1205.
[27] Trope Y, Liberman N. Construal-level theory of psychological distance. *Psychol Rev*. 2010;117(2):440-463. doi:10.1037/a0018963
[28] Kearney MW. Rtweet: collecting and analyzing twitter data. *J Open Source Software*. 2019;4(42):1829.
[29] Silge, J., & Robinson, D. (2017). *Text mining with R: A tidy approach*. "O'Reilly Media, Inc." https://www.tidytextmining.com
[30] Russell JA. Core affect and the psychological construction of emotion. *Psychol Rev*. 2003;110(1):145-172.
[31] Kensinger EA. Remembering emotional experiences: the contribution of valence and arousal. *Rev Neurosci*. 2004;15(4):241-252.
[32] Siegle, G. J. (1994). The balanced affective word list creation program. http://www.sci.sdsu.edu/CAL/wordlist/
[33] DeRose, S. (2005). The Compass DeRose Guide to Emotion Words. http://www.derose.net/steve/resources/emotionwords/ewords.html
[34] Dixon, S. (2022). Countries with most twitter users 2022|Statista: https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/.
[35] Clement, J. (2022). Twitter desktop traffic share 2022|Statista. https://www.statista.com/statistics/261402/distribution-of-twitter-traffic-by-country/.
[36] Mylonas, D., Purver, M., Sadrzadeh, M., MacDonald, L., & Griffin, L. D. (2015). The Use of English Color Terms in Big Data. AIC 2015 Tokyo, Color and Image: Midterm Meeting of the International Color Association, Tokyo, Japan.

AUTHOR BIOGRAPHIES

Boshuo Guo is Lecturer of Digital Fashion Marketing at School of Design, The University of Leeds, Leeds, UK.

Stephen Westland is Professor of Color Science, School of Design, The University of Leeds, Leeds, UK.

Peihua Lai is PhD Student at School of Design, The University of Leeds, Leeds, UK.

How to cite this article: Guo B, Westland S, Lai P. Sentiment analysis based on frequency of color names on social media. *Color Res Appl*. 2023; 48(2):243-252. doi:10.1002/col.22837

APPENDIX

Ten most frequently used words and their sentiment scores for each of the colors

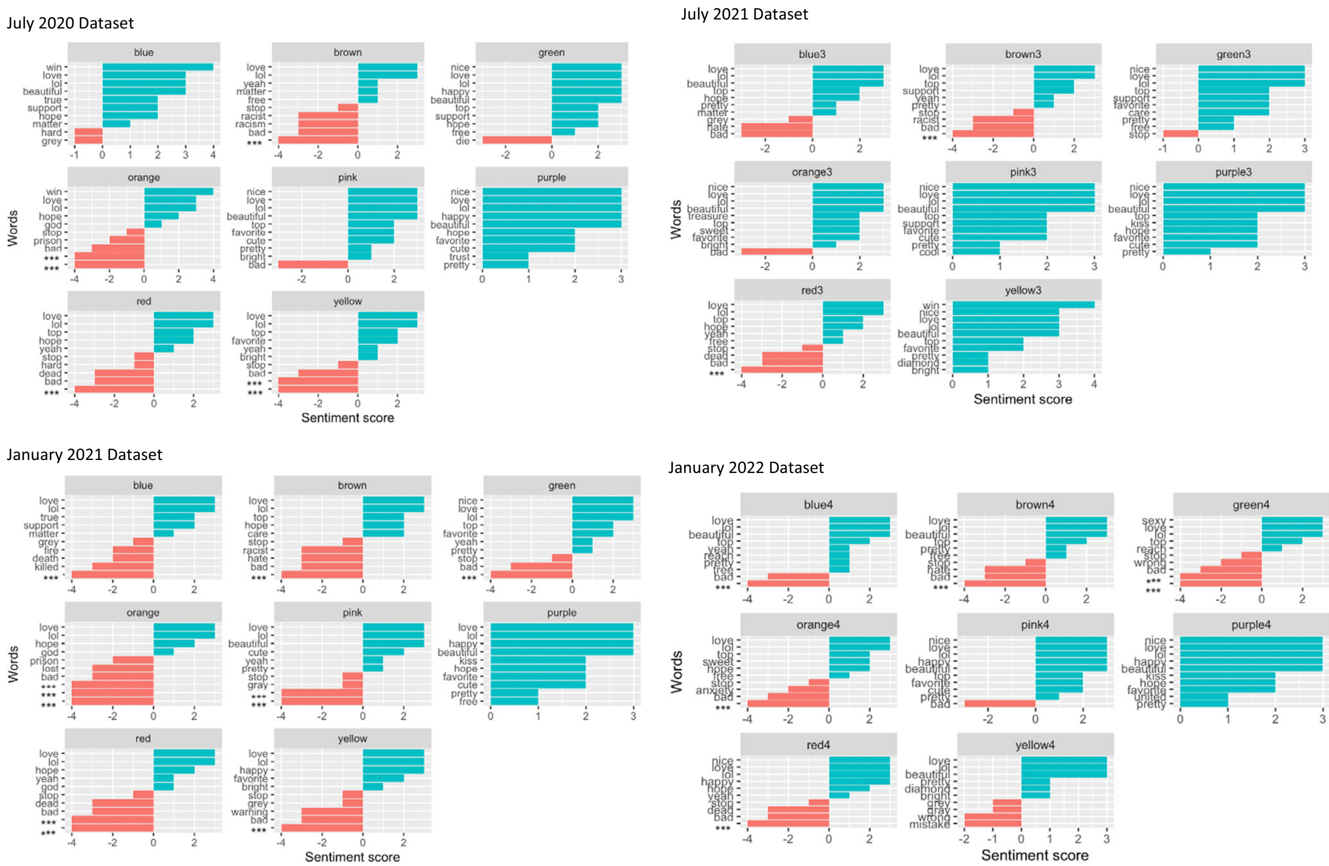

FIGURE A1 show the 10 most frequently used words, although the frequency of each word is not displayed on the figures. In these figures, we use green bars to indicate positive sentiment scores, whilst red bars are used for negative sentiment score. Since words in these figures are retrieved from the original Twitter posts, there were some impolite words that may cause offence. We replace these impolite words with ***