

Self-Representation on Twitter Using Emoji Skin Color Modifiers

Alexander Robertson, Walid Magdy, Sharon Goldwater

Institute for Language, Computation and Cognition
 School of Informatics, University of Edinburgh
 alexander.robertson@ed.ac.uk, {wmagdy, sgwater}@inf.ed.ac.uk

Abstract

Since 2015, it has been possible to modify certain emoji with a skin tone. The five different skin tones were introduced with the aim of representing more human diversity, but some commentators feared they might be used as a way to negatively represent other users/groups. This paper presents a quantitative analysis of the use of skin tone modifiers on emoji on Twitter, showing that users with darker-skinned profile photos employ them more often than users with lighter-skinned profile photos, and the vast majority of skin tone usage matches the color of a user’s profile photo—i.e., tones represent the self, rather than the other. In the few cases where users do use opposite-toned emoji, we find no evidence of negative racial sentiment. Thus, the introduction of skin tones seems to have met the goal of better representing human diversity.

Introduction

Emoji—icons used to represent emotions, ideas, or objects—became a formally recognized component of the Unicode Standard in 2010. At that time, all emoji depicting humans were rendered with the same yellow skin tone. However, in 2015, special modifier codes were introduced to allow users to change the skin tone of certain emoji. Unicode Technical Standard #51 (Davis and Edberg 2014), justified this on the grounds that “people all over the world want to have emoji that reflect more human diversity.” Nevertheless, the change was not uncontroversial. Although skin tone modifiers can be used to make emoji more personal to reflect the identity of the user, in principle they can also be used to represent others. This possibility sparked fears in the mainstream media that users might engage in “digital blackface”, negatively using dark emoji for entertainment purposes (Princewill 2017; Dickey 2017)—if people would even use them at all (Zimmerman 2015). But despite this media and popular attention, and considerable academic investigation of emoji in general, we are not aware of any quantitative studies of how people actually use skin tone modifiers.

In this paper, we present an analysis of how skin tone modifiers are used in emoji on Twitter. In a sample of 0.6 billion tweets, we find that 42% of the 13 million tone-modifiable emoji (TME) tweeted in 2017 included skin tone modifiers (henceforth, we refer to these modified TME as TME+, in

Copyright © 2018, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.



Figure 1: The current 102 tone-modifiable emoji. The first five rows are modified by a skin tone (shown in the left-most column), and the final row contains unmodified emoji.

contrast to unmodified TME-) and the majority of these TME+ were light colors. By annotating the profile photos of 4,099 users and analyzing their tweet histories, we show that the overall prevalence of lighter skin tone modifiers is likely due to the prevalence of lighter-skinned users. Indeed, users with darker-skinned profile photos are more likely to use TME, and more likely to modify them. Moreover, the vast majority of TME+ are similar in tone to the user’s profile photo, suggesting that TME+ are used overwhelmingly for self-representation. Finally, we analyze the sentiment of the small proportion of tweets containing opposite-toned emoji, and find that positive tweets outweigh negative. Overall, our findings suggest that the Unicode Consortium has successfully met a real demand for diversity within emoji, with few negative consequences.

Emoji and Skin Tones

Recent years have seen significant academic interest in emoji. Studies show that emotional interpretation of some emoji is subject to ambiguity (Miller et al. 2017), that addition of emoji to messages (particularly neutral ones) can convey positive/negative tone and sentiment (Hu et al. 2017), and that popularity differences between emoji may be due to their relation to popular words (Ai et al. 2017). Resources commonly available for traditional language such as sense inventories have been compiled (Wijeratne et al. 2017) and emoji usage has been linked to personality traits (Marengo, Giannotta, and Settanni 2017). But despite three years since their introduction, TME have been unstudied until now.

Not all emoji can have a skin tone. The traditional emoji

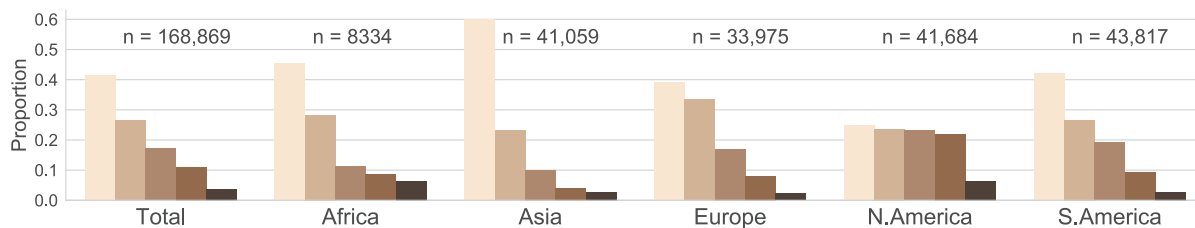


Figure 2: Proportion of TME+ using each skin tone, by geographic region.

faces (e.g. 🍌) cannot be modified. The current Unicode Standard defines 102 TME, some of which can be combined to create compound emoji (e.g. 🍌👤) which can also be modified for skin color. Five skin tones, based on the six tones of the Fitzpatrick phototyping scale (Fitzpatrick 1988), are available. We will refer to the five-scale tones as T1 (the lightest) to T5 (the darkest). Figure 1 illustrates the 102 available TME, showing how the tones affect them.

Emoji are rendered differently across platforms, which can cause ambiguity in semantic interpretation—e.g., writers/readers may perceive the expressed action or emotion differently (Miller et al. 2017). Since we are only concerned with skin color, we do not expect these ambiguities to be an issue in our study. However, there is slight variation in the rendering of specific skin tones across platforms. We therefore cannot be sure of the precise tonal values observed by a user, although all platforms have exactly 5-scale tones, where T1 is lighter than T2, T2 lighter than T3, and so forth.

Skin Tone Usage, Overall and By Region

For our initial analysis we used the Twitter API (1% sample) to create Dataset A, from 1.04 billion tweets made between February and December 2017. This dataset contains 595m original (non-retweet) tweets made by 769m unique users. Of these, 83.3m (14%) contain at least one emoji and 13.1m (2.2%) contain at least one TME. 5.5m (42%) of TME were TME+. Lighter tones dominate: T1 and T2 account for 68% of all TME+.

To gain a global view of TME usage by location, we used the Google Maps API to validate the location field of the authors of the 5.5m TME+ tweets¹. We located 169k users, then grouped them on the basis of broad geographic region (figure 2). Most regions follow the overall trend of using far more light tones. The exception is North America, where the distribution of tones is more uniform—though T5 still has relatively little usage, and is the least common tone worldwide. This raises the question of why lighter TME+ are used so much more than darker ones in all regions outside North America. These results would seem to contradict the hypothesis that TME+ are used for self-representation, since we might naïvely expect more dark-skinned users in Africa and Asia, whereas the TME+ being used are mainly light. However, many countries (particularly in Africa) have lower levels

¹Though this field is not necessarily reliable (users may travel or give a false location), our aim was to get a broad overview before the main analysis.

of internet access compared to highly developed countries (such as those in North America)², and access to the internet or modern devices can be even more limited for some ethnic groups due to political oppression (Weidmann et al. 2016) or poverty. So, it could simply be that in most regions, lighter-skinned internet users predominate.

Another possible hypothesis for North America’s more uniform TME+ distribution (although the demographics do not have similar distributions) is that people use TME+ to refer to both others and themselves. We investigate this hypothesis in the next section, looking more closely at TME usage on Twitter by comparing users’ actual skin tone to that seen in their TME+.

Skin Tone Modifiers for Self-Identification

Since our goal is to better understand how TME relate to user identity, we randomly selected 10k users from Dataset A, subject to the condition that they had used at least one TME. Although we cannot perfectly determine the skin color of these users, we estimate it from their profile photos. We assume that photos of a single individual are self-portraits of the user. This might not be true in all cases, since users can use photos of others. We assume this practice is rare, with a negligible effect on our analysis. However, our annotation process was designed with this possibility in mind.

We collected the Twitter profile pictures of the selected 10k users. These pictures were then annotated using Crowdflower³. We instructed annotators to label as “invalid” images that might not represent the Twitter user. “Invalid” images include: non-human (e.g. animals, nature), celebrities, group photos, gray-scale images. Instructions were given beforehand, with examples of valid/invalid photos provided. If a picture was annotated as “valid”, annotators compared it to an array of five TME+, as shown in Figure 3. We selected the example emoji for its overall lack of distracting features (e.g. little hair covering the face) and consistent appearance (all versions have the same hair/eye color).

Each photo was annotated by at least three annotators. A set of 50 pre-annotated pictures was used to control the quality of the annotators. Only workers achieving over 90% accuracy on the screening pictures were accepted.

After the annotation process, we retained only those pictures annotated as valid where at least two annotators agreed

²World Bank, World Telecommunication Report, <http://data.worldbank.org/indicator/IT.NET.BBND.P2>

³www.crowdflower.com

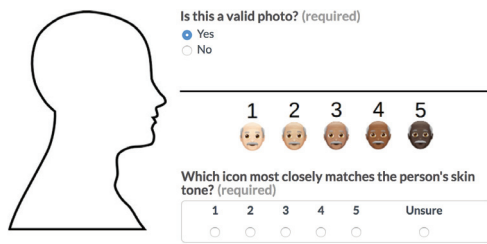


Figure 3: Example of Crowdflower annotation task.

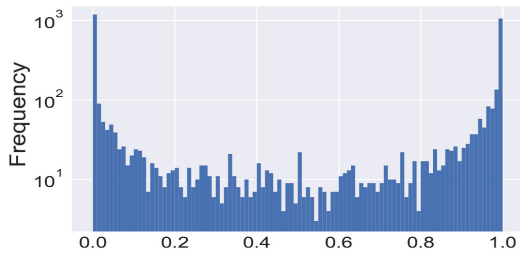


Figure 4: Number of users (y-axis, log scale) who use TME+ at a given proportion (x-axis) of the time.

on the same skin tone. The result was 4,099 annotated pictures. The inter-annotator agreements for these pictures were 98% for validity, 88% for user skin tone.

TME Usage Analysis

We collected the historical tweets of the 4,099 annotated users using the Twitter API, to create Dataset B. This contains the tweets of these users from January 2018 back to April 2015—the time when TME were enabled. 5.86m original (non-retweet) tweets were collected in total. 2.46m tweets contained emoji, of which 479k contained TME.

Before looking at the relationship between a user’s skin tone and their use of TME, we first examined the overall usage of TME by user. The histogram in Figure 4 shows that, of the Dataset B users, roughly a quarter only ever use TME-, with about the same proportion always using TME+. The remaining users tend towards one end or the other.

Next, we looked at how TME+ usage varies across users according to their annotated skin color. Figure 5 shows the proportion of users in each of the five skin color groups who tweeted at least one TME+. A clear trend is visible, where users with darker skin are more likely to use TME+ than those with lighter skin. In fact, 80+% of users with skin tone 4 and 5 use TME+ in their tweets, which shows the importance of this feature to them.

However, not all TME are TME+. Thus, we examined users who use TME and computed the proportion of those tweets which are TME+. Figure 6 shows the percentage of TME+ vs TME- for each user group. Again, the same trend appears: darker skinned users tend to modify the skin tone of TME in their tweets much more than users with lighter skin. For instance, 82% of TME tweets by users with T4 skin change the skin tone of the emoji compared to only 52% of the white (T1 skin) users. This emphasizes how users with

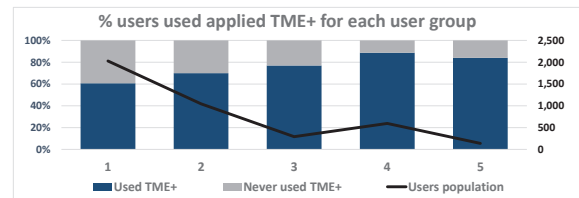


Figure 5: Percentage of users in each skin tone group (1=lightest, 5=darkest) who used TME+ at least once. Black line (right-hand axis): number of users per group.

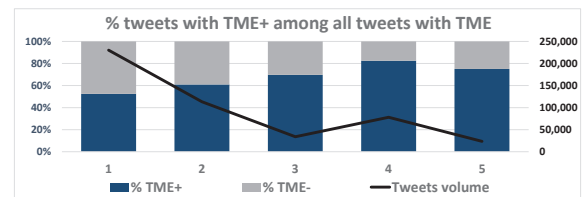


Figure 6: Percentage of tweets with TME+ made by users in each skin tone group, out of all tweets with TME. Black line (right-hand axis): number of tweets per group.

darker skin are keen to modify the appearance of their emoji, presumably in order to better reflect their own identity.

TME+ for Self-Reference vs Referencing Others

Finally, we looked at whether users’ choice of modifiers reflects their own skin color or others. For each tone user, we determined the TME+ tone used most often by that user. Figure 7 shows a heatmap of the distribution of the most common skin tones used by each user group in Dataset B. Overall, our results suggest that by far the most common scenario is for users to choose skin tone modifiers that roughly match their own complexion, accounting for the variability in rendering of these tones across different platforms. There seems to be some tendency for users to choose modifiers towards the lighter end of the available spectrum; for example, users with skin tone 2 mostly select skin tone 1 in their emoji, and users of skin tone 5 mostly select skin tone 4 in their emoji.

The (Non)-Prevalence of Digital Blackface

The previous analysis shows that for nearly all users, their *most common* choice of skin tone matches their own skin. However, it still may be the case that users occasionally choose emoji of a different tone. We looked at how often this occurs, and whether there is any noticeable negative sentiment associated with this behavior. For this analysis, we aggregated Dataset B users into two groups based on skin tone: tones 4/5 (dark) and tones 1/2 (light). First, we analysed the tweets of each group that use TME+ with the skin tone of the other group. These comprise 1229 tweets for the dark-skinned users and 1,729 tweets for the light-skinned users—a very small proportion of each group’s tweets: 0.14% (light group) and 0.24% (dark group). Next, we examined the opposite-tone tweets in English and computed their sentiment using Sentistrength (Thelwall, Buckley, and Paltoglou 2012),

1	0.6	0.3	0.08	0.01	0.007
2	0.5	0.2	0.2	0.03	0.008
3	0.2	0.1	0.5	0.2	0.02
4	0.03	0.01	0.3	0.6	0.04
5	0.03	0.009	0.2	0.7	0.1
	1	2	3	4	5
	Emoji tone most often used				

Figure 7: Most common emoji skin tone used vs perceived user skin tone, normalized by row sum.

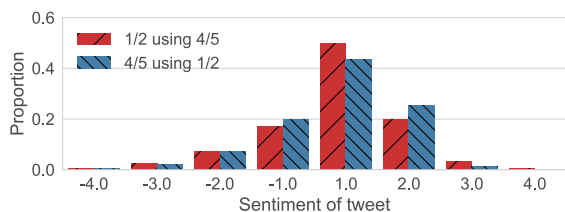


Figure 8: Distribution of non-neutral sentiment for English tweets containing lighter/darker TME+ but written by users having darker/lighter skin.

which is designed specifically for analyzing short online messages such as Twitter posts. The majority of these tweets were neutral (almost 50% of the tweets of both groups). The distribution of the non-neutral sentiment of tweets for both groups are shown in Figure 8, where 4 is the most positive and -4 is the most negative⁴. As shown, the distributions for both groups are almost the same, and positive tweets outnumber negative ones. Inspection of tweets with negative sentiment revealed tweets on generally negative topics, rather than anything specifically about race. Overall, we find no evidence to justify fears of widespread “digital blackface” or its black-against-white counterpart.

Conclusion

In this paper we presented the first quantitative study on the usage of tone-modifiable emoji (TME) on Twitter. We showed that although lighter-toned emoji are more common overall, different populations of users (based on the skin color of their profile picture) show markedly different characteristics in how they use TME and to what extent. The overall picture is one where users take advantage of emoji skin tone modifiers to represent an important aspect of their identity, and do so differentially depending on their own skin tone: compared to light-skinned users (the majority on Twitter), a higher proportion of dark-skinned users use skin tone modi-

⁴Sentiment was rated on a -5:+5 scale but no tweets at the extremes of this range were found in this subset of the data.

fiers, and they use them more frequently. Here we grouped all geographic regions together in looking at user skin tone, but an interesting question for future work will be to examine the relationship between a user’s majority/minority status within their real-life community (i.e., their geographic region) and their use of skin tone modifiers.

Acknowledgements

This work was supported in part by the EPSRC Centre for Doctoral Training in Data Science, funded by the UK Engineering and Physical Sciences Research Council (grant EP/L016427/1) and the University of Edinburgh.

References

- Ai, W.; Lu, X.; Liu, X.; Wang, N.; Huang, G.; and Mei, Q. 2017. *Untangling emoji popularity through semantic embeddings*. AAAI press.
- Davis, M., and Edberg, P. 2014. Proposed Draft Unicode Technical Report: Unicode Emoji. Technical Report 51, Unicode Consortium.
- Dickey, M. 2017. Thoughts on white people using dark-skinned emoji. *TechCrunch*.
- Fitzpatrick, T. 1988. The validity and practicality of sun-reactive skin types I through VI. *Archives of Dermatology* 124(6):869–871.
- Hu, T.; Guo, H.; Sun, H.; Nguyen, T. T.; and Luo, J. 2017. Spice up your chat: The intentions and sentiment effects of using emoji. *CoRR* abs/1703.02860.
- Marengo, D.; Giannotta, F.; and Settanni, M. 2017. Assessing personality using emoji: An exploratory study. *Personality and Individual Differences* 112:74 – 78.
- Miller, H.; Kluser, D.; Thebault-Spieker, J.; Terveen, L.; and Hecht, B. 2017. *Understanding emoji ambiguity in context: The role of text in emoji-related miscommunication*. AAAI Press. 152–161.
- Princewill, V. 2017. Is it OK to use black emojis and gifs? *BBC News*.
- Thelwall, M.; Buckley, K.; and Paltoglou, G. 2012. Sentiment strength detection for the social web. *Journal of the Association for Information Science and Technology* 63(1):163–173.
- Weidmann, N. B.; Benitez-Baleato, S.; Hunziker, P.; Glatz, E.; and Dimitropoulos, X. 2016. Digital discrimination: Political bias in internet service provision across ethnic groups. *Science* 353(6304):1151–1155.
- Wijeratne, S.; Balasuriya, L.; Sheth, A. P.; and Doran, D. 2017. EmojiNet: An open service and API for emoji sense discovery. In *ICWSM*, 437–447.
- Zimmerman, J. 2015. Racially diverse emoji are a nice idea. But will anyone use them? *The Guardian*.