THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Emoji Skin Tone Modifiers

OPEN ACCESS

Emoji Skin Tone Modifiers: Analyzing Variation in Usage on Social Media

ALEXANDER ROBERTSON, School of Informatics, University of Edinburgh, Scotland
WALID MAGDY, School of Informatics, University of Edinburgh, Scotland
SHARON GOLDWATER, School of Informatics, University of Edinburgh, Scotland

Emoji are widely used in computer-mediated communication to express concepts and emotions. Skin tone modifiers were added in 2015 with the hope of better representing user diversity, and indeed recent work has shown that these modifiers are especially popular amongst darker-skinned users, who are a minority on Twitter. Previous work also showed that the vast majority of tone-modified emoji have a tone similar to the user's own skin tone, suggesting that self representation is a major factor in tone use. In this paper, we first show that the basic finding (users mainly choose a tone that is similar to their own skin tone) generalizes to different sub-populations of users, including users from majority-Black regions. We then extend the analysis of tone use to quantify and examine cases where users *modulate* their tones: that is, for a particular emoji, they choose either to use a different tone than their usual one, or no tone at all (after having previously used one). We show that even though these uses constitute only a small proportion of emoji usage, many instances are readily classifiable as ways of representing *other* people. The evidence we present is therefore crucial in working towards a broader understanding of the connection between emoji and identity expression online. We also offer explanations for why the darkest emoji skin tones are not used, by examining aspects of their design which make them less suited to self-representational usage. This highlights the need for careful consideration of both design and human diversity when creating emoji. Moreover, despite early fears in the media, we find little evidence of negative usage even when tones are used in a non-self-representational manner. In sum, our findings lend even more support to the highly positive role that emoji and skin tone modifiers play in identity and expression in computer-mediated communication.

CCS Concepts: • **Applied computing** → **Law, social and behavioral sciences**; • **Human-centered computing** → *Human computer interaction (HCI)*;

Additional Key Words and Phrases: emoji, skin tone, social media, identity, self-representation, representation

1 INTRODUCTION

Human diversity and expression of identity go hand in hand: without the former, the latter would likely be unnecessary. Online identity, especially how internet users express themselves through text, is well-studied, with one key question being how online personas reflect or are influenced by offline reality. In multi-user online games, Chester and Bretherton [2007] found that users enhanced

Authors' addresses: Alexander Robertson, School of Informatics, University of Edinburgh, Edinburgh, Scotland, alexander.robertson@ed.ac.uk; Walid Magdy, School of Informatics, University of Edinburgh, Edinburgh, Scotland, wmagdy@inf.ed.ac.uk; Sharon Goldwater, School of Informatics, University of Edinburgh, Edinburgh, Scotland, sgwater@inf.ed.ac.uk.
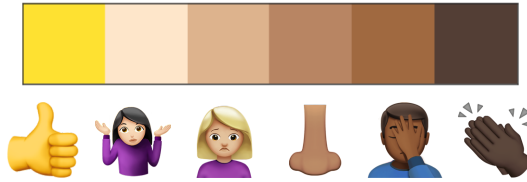
Fig. 1. The original yellow emoji color and the five emoji skin tones (as viewed on Apple devices), added in 2015. We refer to the unmodified yellow color as tone 0 and the non-yellow tones as 1 to 5, from lightest to darkest. Example modified emoji are provided below: thumbs up, shrug, frown, nose, face-palm, clapping hands.

existing traits rather than create entirely new ones, and that, despite the setting being far removed from reality, users wished to present their "real" self to some extent. In the context of social media, Zhao et al. [2008] also found that users tended to present enhanced, idealistic versions of themselves. However, identity can also be highly performative and at odds with reality, as Ringrose [2011] found in a study of young people on Bebo and their expression of sexual identity through text and photos. And not all users engage in identity expression in the same way. Grasmuck et al. [2009] found that, compared to their Vietnamese and White peers, African American and Latinx Facebook users at US colleges "intensively invested" in modes of identity construction such as displaying photos and explicitly stating their tastes and interests.

The aforementioned studies have focused on text and photos as means of identity expression. Recently, a third means of expression has emerged—emoji. Since April 2015, it has been possible to alter the default yellow appearance of certain emoji by applying one of five skin tone modifiers [Davis and Edberg 2014]. Of the roughly 1200 base emoji such as 👋 or 😍, 106 can be combined with a skin tone modifier (Figure 1). These tone-modifiable emoji (TME) typically represent humans, and include images of body parts (👌, 💪), activities (🏌, 🏊), facial expressions (🤦), and body language (💁, 🙅). The introduction of skin tone modifiers was justified as a way to "reflect more human diversity" [Davis and Edberg 2014]. Unlike text or photos, then, their sole intended purpose is to aid in the expression of one's skin color. Toned emoji therefore offer a completely unique means by which to study the ethnic and racial dimensions of online identity expression.

In an initial study of emoji tone usage, we argued that emoji skin tones not only reflect human diversity, but encourage the expression of it [Robertson et al. 2018]. In particular, we analyzed the TME usage of approximately 4000 random Twitter users (with profile photos from which to determine user skin tone) who had used at least one TME. We showed that users' most common emoji tone was one that reflected their own skin tone, and that users with darker skin were more likely to use tone modifiers than those with lighter skin. We argued based on these results that skin tone modifiers are largely used for self-representation, and that this form of self-representation is particularly important for users with darker skin. However, the study left a number of questions unanswered, which we address here.

First, do the results generalize to different sub-populations of Twitter users? In particular, do users from predominantly dark-skinned parts of the world use tone modifiers similarly to users from areas where lighter skin is the norm? The answer to this question could shed light on how users interpret the default yellow emoji. If it is seen as simply a neutral or unspecified human, then we might expect to find higher levels of modification amongst users whose own skin is *different* from that of the average person in their region; i.e., light-skinned users where dark skin is the norm, and dark-skinned users where light skin is the norm. On the other hand, if the yellow is in

fact interpreted as a light color, then dark-skinned users might be expected to modify it more often regardless of their surroundings.

To answer this question, we follow a similar procedure to that of Robertson et al., but use data from approximately 6000 random users as well as from three distinct sub-populations with very different demographics: roughly 4500 users each from London, New York, and Johannesburg. We show that usage patterns are very similar across all four groups, suggesting that dark-skinned users in all regions may interpret the yellow emoji as a lighter-skinned human and prefer emoji that more closely represent their own skin tone.

Second, although users' most common tone is typically *similar* to their own skin tone (as judged by human annotators), why is it often not the *closest* tone? In particular, users annotated as having skin tone 2 (the second-lightest tone available) very often use tones 1 or 3 (as seen in Figure 1), while users with skin tone 5 (the darkest tone) most often use tone 4. Our analysis here suggests that the first effect is due to an aspect of self-representation not discussed previously (hair color), while the second is due to the visual affordances of the darkest tone, which on most platforms tends to obscure details of the emoji image.

Finally, while self-representation appears to be the typical use of tone, users do also choose other tone modifiers in some cases. What can be learned from examining this aspect of user behavior? This paper provides the first analysis we know of to look at these alternative uses of tone. We quantify the extent of *tone modulation* (when a user changes the tone of a particular emoji from one use to the next) and provide a qualitative analysis of the small proportion of English language tweets where users switch to a tone that is clearly different from their own skin tone, or switch between toned and non-toned (yellow) versions of the same TME.

We find that tone modulation is very rare: in our data sets, users change the tone of an emoji compared to its previous usage (including turning it off or on) less than 4% of the time. Of the 338 cases we analyzed where the switch was from one tone to another, about two thirds seem to be references to other users, with other cases referring to groups, self, or specific aspects of the emoji. Switching off an emoji tone where it was previously used is even less common than switching from one tone to another. Of the 60 cases we examined and were able to interpret, nearly a third were cases where a user directly referred to another user and the skin tones of the two users' profile photos differed. On the one hand, these analyses show that skin tone modifiers are used for referencing others as well as the self, though apparently less frequently. On the other hand, they show that users are sometimes reluctant to "choose" an identity for another user, and instead default to the less identity-linked yellow option.

To summarize, the main contributions of this paper are:

- We replicate and extend the main findings of Robertson et al. [2018], showing that around the world Twitter users select emoji skin tone modifiers that closely match their real-life skin tone, and that usage of tone modifiers is higher amongst darker-skinned users, regardless of the demographics of their home location (Section 4.1).
- We demonstrate that the distribution of available emoji tones does not provide even coverage within the color space and the darkest tone obscures fine detail in the final rendered emoji, which accounts for why users with the darkest skin tones are observed to select a slightly lighter emoji tone (Section 4.2).
- We provide the first quantitative and qualitative analysis of *tone modulation*, where in specific cases users actively disable their use of emoji tone modifiers or use a tone different to their own (Section 5). We argue that many of these events reflect users' choices about how to represent other users.

Taken together, these results further our understanding of how skin-toned emoji are used to represent both self and others. Our results should be of interest to designers as well as social scientists, since we found that the availability of darker tones is clearly important for users, yet the visual affordances of the darkest tone seem to inhibit its use. Finally, by highlighting the phenomenon of tone modulation and providing an initial qualitative analysis of it, we hope to stimulate further research into aspects of emoji which are not immediately apparent from looking only at the most common usages.

2 BACKGROUND AND RELATED WORK

Although emoji have been a feature of computer-mediated communication since their introduction in Japan in 1999 [Blagdon 2013], their widespread use was not established for another decade. Important milestones for this include incorporation into the Unicode standard in 2010 and inclusion in iOS 5.0 in 2011 and Android 4.4 in 2013. In recent years there has been a parallel growth of research into emoji, covering a variety of topics. This section first presents a very brief overview of the current theoretical and applied research on emoji in social media. We then discuss prior work on how ethnic identity is expressed in the real world and on social media, and conclude with how these aspects relate to emoji.

2.1 Emoji and Social Media

It has been shown that emoji are used for various purposes, such as establishing the emotional tone of a message [Kaye et al. 2016] and expressing or strengthening particular sentiments [Hu et al. 2017]. Although some emoji can be used as replacements for the objects they literally represent, not all usage is so direct, with personalized interpretations (based on the visual affordances of the images) between individuals and small groups being possible [Wiseman and Gould 2018]. Paralinguistic and pro-social usage has also been studied, with emoji being especially common in online statements of solidarity following crisis events [Santhanam et al. 2018].

Complementing research into emoji usage is a significant body of work on how emoji are interpreted by viewers. Much of this has been on sentiment and affectivity, showing that most emoji have positive interpretations and that people generally agree on the extent of a given emoji's affectivity—a finding that is consistent across multiple languages [Novak et al. 2015]. Differences in how different platforms render emoji can, however, cause confusion because these non-identical forms of the same underlying emoji can have very different interpretations, whether in isolation [Tigwell and Flatla 2016] or with textual context [Miller et al. 2017].

2.2 Self-Representation on Social Media

The most prominent framework of self-representation is that of dramaturgical analysis, following Goffman [1959]. This characterizes in-person interaction as a series of theatrical metaphors. People are actors, using props and costumes, performing roles in a particular manner in front of an audience. These roles are often pre-established by society and the audience has expectations regarding the manner of their performance and the props and costumes that should be involved. This framework has been adapted to the setting of social networks and media [Hogan 2010; Semaan et al. 2017; Wood et al. 2016]. One particularly salient aspect of this adaptation is the notion of "context collapse" [boyd 2007; Wesch 2009] in social media, whereby online interactions between users can potentially be witnessed by the entire world. Moreover, they are often persistent and searchable. Together, these properties mean that it is no longer possible to know precisely who the audience is. As a result, authentic self-expression online can be a treacherous balancing act.

Prior studies, all within the Goffman framework, looked at how users of social media navigate these issues and express particular aspects of their identity online using modes of expression which

pre-date emoji, such as text and photos. For example, Lebel and Danylchuk [2012] and Bigsby et al. [2019] analyzed the words (but not the emoji) in tweets from professional and college athletes, respectively, to see how these athletes craft their identities on Twitter.

Kapidzic and Herring [2015] examined profile photos used on a social networking site and found significant differences in gaze, posture, dress and camera distance between races and sexes. Looking more generally at how Twitter users perform in front of other users in the context of trending topics, Papacharissi [2012] highlighted the importance of play (e.g. through syntactic flexibility and use of literary conventions) when performing the self online.

Building on these studies, our work provides the first insight into how users employ the affordances of emoji as an additional way to shape their identities in the complex online social world.

2.3 Self-Representation of Racial and Ethnic identity beyond Social Media

Skin color and ethnic/racial identity may seem closely connected, but there is little work explicitly connecting the two in terms of self-representation. Skin color on its own has been studied intensely, with a large body of work on attitudes towards one's own skin and that of others: see studies on the use of whitening cosmetics in Jamaica [Charles 2010] and Japan [Ashikari 2005], or the impact of skin color on the resettlement experiences of refugees Australia [Colic-Peisker 2005]. More holistically, Phinney and Ong [2007] provides a thorough overview of the components of ethnic identity, from which skin color is conspicuously absent—the focus is on factors such as self-labeling, exploration, values and beliefs. The labels that ethnic groups use in relation to themselves are especially important, as a means of self-determination and expressing in-group unity. For example, Larkey et al. [1993] studied the ethnic labels Black Americans used to describe themselves, finding that a variety of terms were used. Each was attached to specific senses of identity, heritage, pride and kinship. A change in preference of some terms over others was connected to self-determination and the power to self-label rather than accepting imposed labels. Looking at younger members of ethnic groups, specifically Asian-/Black-/Mexican-American college students, Phinney and Alipuria [1990] argue that ethnic identity is a major part of identity development overall, relative to a control group of White-American college students. Ethnic identity is also seen as more important than political or religious facets, though not occupation or sex role. Working from this perspective, the connection between skin color and expression of identity merits investigation.

2.4 Emoji Modifiers, Self-Representation and Social Media

The introduction of TME in 2015 was met with considerable media and public interest, but also concern that the modifiers would never be used because the number of new emoji being introduced was out-pacing demand for them [Zimmerman 2015], or that they would be used abusively due to their innate link to racial and ethnic identity [Dickey 2017]. However, recent studies suggest that such concerns were unfounded. First, tone modifiers have proven very popular: for example, our previous work [Robertson et al. 2018] showed that, in a corpus of 0.6 billion tweets, 42% of 13 million TME instances were tone-modified. Also, looking specifically at hand-related TME in 22 million tweets geolocated in the USA, Barbieri and Camacho-Collados [2018] found 70% of 414,000 total TME were tone-modified.

Furthermore, by manually annotating a random sample of over 4000 Twitter users' profiles for skin tone, we found that users rarely employ a tone modifier that significantly differs from their own skin tone [Robertson et al. 2018]. The sentiment of tweets containing such mismatched tones was generally positive, while the extremely few instances of negative tweets with mismatched tones were not readily classifiable as racially-motivated abuse. This led us to conclude that tone modifiers are overwhelmingly used for self-representation, in line with the Unicode Consortium's

goal of providing a means for emoji to reflect human diversity [Davis and Edberg 2014]. These emoji therefore represent a major means by which the skin color aspect of one's identity can be succinctly expressed on social media, perhaps similar to the role that ethnic labels [Larkey et al. 1993] have traditionally played. The current study expands on our previous work, providing more detailed investigation to further elucidate the role of skin tone modifiers in representing both the self and others.

3 DATASETS

We constructed our datasets using the Twitter Streaming API (1% sample) and assigned randomly selected users to one of three groups based on their location, until each group contained 10,000 users. A detailed description of the data collection and annotation process is provided in Appendix A. Unlike Robertson et al. [2018], we do not look for users who have used at least one TME—this could result in a bias, where the data does not afford any insights into those users who never use TME. The locations we focus on are Johannesburg, London and New York City. Users were considered to be based in a location if the self-provided location on their profile matched one of the three targets or, in the case of London and New York City, a borough such as Camden or The Bronx.

In addition, we selected a fourth group of 20,000 random users.[1] Although users in the random group may state their location as Johannesburg, London or NYC, we made sure not to include the *same* user in more than one group. Therefore, the users are unique across all groups.[2]

The motivation for the selected three locations is their demographic composition, each having different proportions of ethnic groups which are likely to have skin tones at different points of the range currently represented through TME. In addition, each location has a large proportion of English speakers (which is useful for our qualitative analysis). Aggregated census data is presented in Table 1. We have collapsed sub-groups into single categories where possible (e.g. White British and White Irish into White, Black African and Black Caribbean into Black) to aid comparison between cities. Johannesburg has a predominately Black population [Statistics South Africa 2011], London predominantly White [Office for National Statistics 2011]. New York City is somewhat more balanced and a large proportion (2.3 million, approximately 28% of the city) of residents identify as Hispanic or Latinx, besides their racial identity [United States Census Bureau 2010].

The Twitter profile photos of these users were then shown to paid annotators on the Figure Eight[3] crowd-sourcing platform, following the method in Robertson et al. [2018]. Annotators were instructed to classify a photo as "invalid" if it contained multiple or no people, if not a full-color photo, or if the photo subject's face was obscured. A "valid" photo passing these checks was then annotated for the skin tone of the subject: annotators compared the photo subject to the five TME skin tone modifiers as shown in Figure 1, which ranges from skin tone 1 (lightest) to skin tone 5 (darkest). Full details on the annotation guidelines are provided in Appendix A. Each profile photo was annotated by three annotators. Only Twitter users where at least two annotators agreed on both photo validity and skin tone were used in this study. Fleiss' Kappa for inter-annotator agreement on photo validity alone was 0.82. When considering only those photos where at least two annotators agreed on both validity and skin tone, Fleiss' Kappa for skin tone was 0.56. The end

---

[1]Users in the random group contained fewer valid photos than those in the Johannesburg, London and NYC groups—approximately 30% for random users compared to 45% for the others. Profile photos are an important part of the annotation process, which we go on to describe. Therefore, we sampled twice as many random users in order to balance out the final number of annotated users in each group.

[2]37% of the random users state no location in their profile, while the percentage of users with locations matching the three target regions is less than 2%.

[3]http://www.figure-eight.com

Table 1. Aggregated racial demographics for three locations, taken from SA, UK and USA census data. Johannesburg is measured in terms of the greater metropolitan area, rather than the city center alone.

	Johannesburg		London		New York City	
	#	%	#	%	#	%
Black	3,389,278	76.42	1,088,600	13.32	2,088,510	25.74
White	544,530	12.28	4,887,500	59.79	3,597,341	44.34
Mixed	247,276	5.58	405,300	4.96	325,901	4.02
Indian/Asian	216,198	4.88	1,511,600	18.49	1,038,388	12.8
Other	37,545	0.85	281,000	3.44	1,062,334	13.1
Total	4,434,827		8,174,000		8,112,474	

Table 2. Number of users per region in our annotated data set (excluding those with private Twitter profiles), broken down by skin tone group and presence/absence of a Twitter profile photo.

	Users per skin tone group (with photo)						
	1	2	3	4	5	**Total**	No photos
Johannesburg	480	136	481	2,995	466	**4,558**	5,065
London	3,346	446	198	423	90	**4,603**	5,093
New York City	2,896	530	246	729	122	**4,523**	9,091
Random	4,085	1,063	287	479	85	**5,999**	11,688
Total	**10,807**	**2,475**	**1,212**	**4,626**	**763**	**19,683**	**26,937**

Table 3. Number of original tweets per user, by region, in our annotated data set (excluding users with private Twitter profiles), for users with and without a Twitter profile photo.

	Users with photos			Users without photos		
	Mean	Std	**Total**	Mean	Std	**Total**
Johannesburg	1,113	1,222	**5,074,264**	617	1,036	**4,932,450**
London	1,628	1,204	**7,493,712**	1,544	1,231	**7,930,530**
New York City	1,429	1,241	**6,456,190**	1,317	1,277	**6,740,760**
Random	1,551	1,408	**9,294,404**	1,984	1,802	**23,771,425**
Total			**28,318,570**			**43,375,165**

result, out of 50,000 profile photos, was 19,683 fully annotated users. General descriptive statistics for users are presented in Table 2.

Finally, in April 2018 we retrieved the most recent tweets for each user (up to 3,200, as per Twitter's API restrictions) and removed any retweets. Since the ability to add tone modifiers to emoji only became possible in April 2015, we did not retrieve tweets made prior to then. We repeated the process again in November 2018, in order to gather more tweets per user. Users who had since set their profile to private were removed from the final dataset. General descriptive statistics for collected tweets are presented in Table 3.

4 EMOJI AND SELF-REPRESENTATION

In this section, we first show that the basic results of Robertson et al. [2018] generalize to the regional sub-populations used here, and then provide explanations for some of the subtler patterns
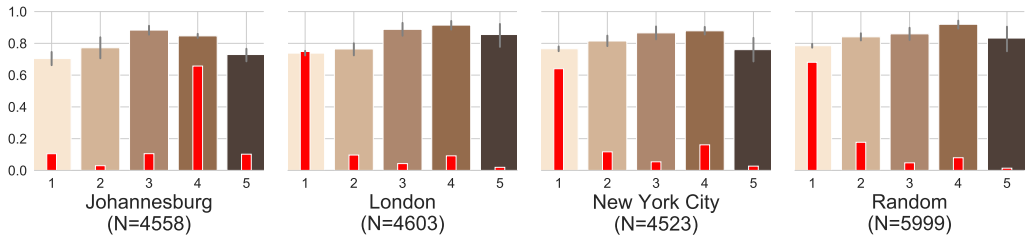
Fig. 2. The proportion of users (with bootstrapped 95% confidence intervals) in each location, per skin tone group (1-5), who have used TME+ at least once. The superimposed red bars show the proportion of users of that skin tone in each location's dataset.

that were noted, but not accounted for, in that paper. We argue that the full pattern of results can be resolved by considering users' desire to express multiple aspects of their visual identity in an emoji while also maintaining its visual interpretability.

4.1 Extending previous findings to distinct sub-populations

One finding of Robertson et al. [2018] was that users with darker skin tones are more likely to have used TME with tone modifiers (henceforth, TME+) than those users with a lighter skin tone, whereas the latter are more likely to use only the unmodified yellow TME (henceforth, TME-). Our results here indicate the same general trend in all three locations and the random user group, as shown in Figure 2. Following Robertson et al.'s grouping of users into lighter-skinned (tones 1 and 2) and darker-skinned (tones 4 and 5), we find the mean proportion of usage is higher in all cases among the darker-skinned users: by 11.2% in Johannesburg, 16.3% in London, 8.8% in NYC, and 11.1% in the random sample. All differences are significant at the $p < 0.001$ level using a chi-squared test of independence on one [skin tone group/TME+ user] contingency table per location. Since these patterns are seen regardless of whether darker skin is the minority (as in London, NYC, random sample) or the majority (as in Johannesburg), this increased usage may be motivated by a user's minority status on Twitter rather than their status in their home location, or simply the fact that the default yellow, rather than being 'unspecified', is perceived as more similar to light-skinned than dark-skinned users. Either way, these usage patterns support the argument that TME+ are a vital means of self-expression for darker-skinned users online, all over the world, no matter where they may live.

That said, we also note that the peak users of TME+ appear to be those with skin tones 3-4, with some drop-off amongst users labeled as tone 5. This pattern was also seen in Robertson et al.'s data, but like our random set, theirs contained very few tone 5 users, making it unclear whether the drop-off was statistically meaningful. Here, we see the same pattern in all 4 location groups, and in the two groups with the largest numbers of tone 4-5 users (Johannesburg and New York), the error bars are small enough to confirm a statistically significant difference in TME+ usage between tone 4 and tone 5 users (chi-squared test of independence $p < 0.001$ in both cases). While these results at first may seem to contradict our main claim, the remainder of this section and the next provide an explanation for why tone 5 users may be using TME+ less than expected.

First, Figure 3 shows that in all four of our sample populations, users' most commonly used skin tone modifier is one that closely matches their own skin tone—i.e., the highest values in the plots are close to the diagonal. However, there are two exceptions to a strict diagonal: First, users with
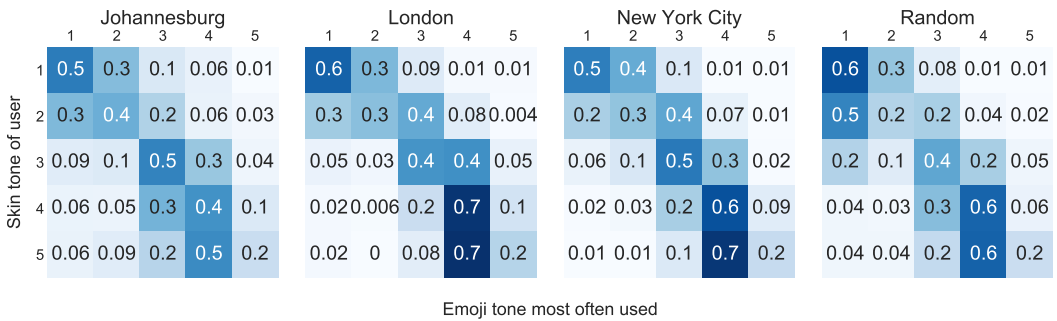
Fig. 3. Most commonly used skin tone modifiers, by user skin tone group. Each row is normalized, showing the proportion of users with that skin tone whose most common emoji tone is the one in the corresponding column. High values near the diagonal show that users in all locations mostly choose an emoji tone that is close to their own actual skin color.

skin tone 2 often (in many cases more commonly) use tones 1 or 3. Second, users with the darkest skin tone most often use the slightly lighter tone 4.

These findings replicate those of Robertson et al. [2018] on the random sample, and show that the results generalize to the other three location groups. However, in that earlier paper, no explanation was provided for the divergences from a strict diagonal pattern, i.e., a direct mapping of emoji skin tone to user skin tone. In the following section, we provide explanations for both of the divergences (at the light and dark end of the scale). These will also help explain why tone 5 users are less likely to have used TME+ than tone 4 users.

4.2 Limitations of current emoji skin tones

Emoji are not rendered identically on all platforms, and even different versions of the same platform may render emoji with slight or major differences. Nevertheless, there are many commonalities. Figure 4 shows examples of hands and faces across the five most commonly used platforms.[4] Apple is synonymous with iOS and Google with Android, with other vendors also providing their own emoji art. As can be seen, facial emoji with skin tone 2 are the only ones which have blond hair. Some users might be avoiding this emoji tone because it does not match their hair, while other users may choose it *because* it matches their hair color, even if it is *not* the closest match for their skin tone. Therefore, the less-than-complete match between user tone and emoji tone may actually reflect additional choices of self-representation, rather than undercutting the hypothesis.

These images also provide a possible explanation for why users with the darkest skin tones don't use TME+ as much as we might expect, and why they overwhelmingly use the lighter tone 4 when they do apply a modifier. In particular, the gradient from light to dark emoji is not uniformly smooth: the difference between 4 and 5 is perceptibly greater than that of any other two adjacent tones. To quantify this observation, we took the most common pixel color for each of the hand emoji, as rendered in each tone by each platform, and determined its hue, saturation and brightness (HSB) values. The HSB model is an alternative to RGB, designed to reflect the way the human visual system produces color sensation [Joblove and Greenberg 1978].

---

[4]Android, iOS and the Twitter Web Client are the top three most commonly used platforms in each of our datasets, as determined by the "source" attribute returned by the Twitter API. The percentage of tweets made via Android/iOS/Twitter are 40/29/9 (Johannesburg), 12/45/19 (London), 11/50/17 (NYC), 26/36/10 (Random). The higher proportion of iOS usage in London and NYC is likely linked to higher gross disposable income in these areas.
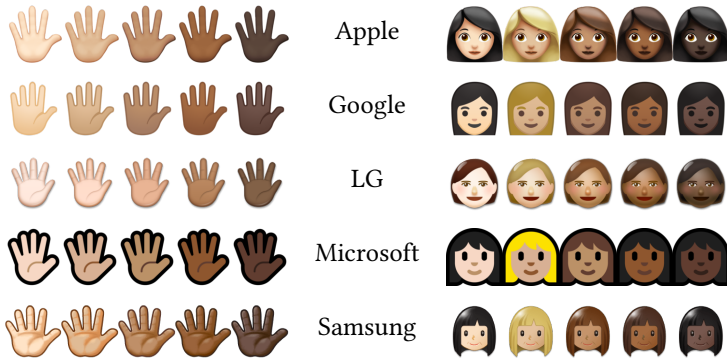
Fig. 4. Differences in tones for five platforms. The magnitude of difference between adjacent tones is significantly greater at the darker end of the range.
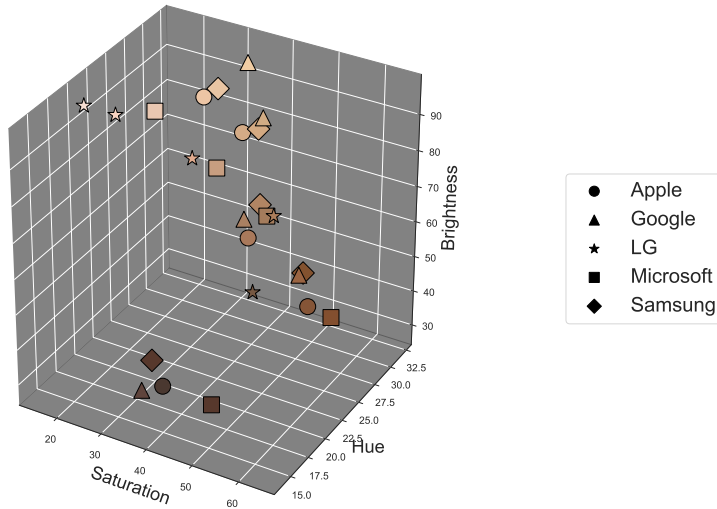

Fig. 5. Hue, saturation and brightness values of predominate pixel color per "hand" TME+, for five platforms. Unlike all other tones, the darkest emoji skin tone is, in general, extremely distant from its nearest neighbors in the tone sequence.

Figure 5 visualizes the HSB values in three dimensions, making it clear that the darkest skin tone for most platforms is an outlier. The exception is tones used on LG phones prior to 2017. It is possible that if we had used the LG tones for annotation, we would have observed a stronger match between users' perceived skin tone and their most commonly-used TME+ tone.

In addition to being more distinct from the other four tones, the darkest skin tone can also make details of the emoji harder to see. As demonstrated in Figure 4, the hair and skin are difficult to separate, while facial features such as the lips and nose become much less perceptible. These issues are exacerbated by the small size of emoji in deployment: compare 😀 and 😀. Users may therefore
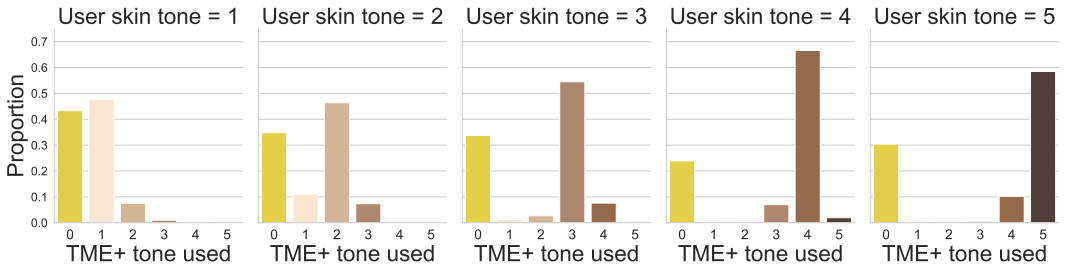
Fig. 6. Evidence for tone modulation. Distributions of tones used in TME+, plus TME- shown in yellow, by users (aggregated by skin tone) whose most common TME+ tone matches their actual skin tone. Users use a variety of tones, besides those which represent their own identity.

avoid tone 5 even if it closely matches their own skin tone, since it does not afford any fine detail. In many cases this means using tone 4, but in some cases it may influence users to avoid applying a modifier at all. This would explain the lower use of modifiers in this group relative to the tone 4 users.

5 TONE MODULATION

The analysis in the previous section focused on which tone is employed by users on social media in the *majority* of their posts. However, these analyses overlook the fact that users actually produce a *variety* of tones in their communications. We illustrate this in Figure 6, which shows the distribution of TME+ tones produced by those users whose majority TME+ tone exactly matches their own skin tone. These are the users in the diagonal cells of Figure 3, aggregated across all datasets. As expected, the most common TME+ tone per group matches the skin tone of that group. Nevertheless, the distribution over all tones indicates that there is a degree of inconsistency in the tones users employ in individual TME+: evidence that users *modulate* their tone production. Furthermore, the second-highest proportion in each group is TME-, which is evidence that users may turn tones on and off completely. This is surprising, since platforms such as iOS and Android remember the last tone used for any given TME, while Twitter's web interface even allows users to globally set the tone for *all* TME.

An alternative explanation for the high proportion of TME- is that individual users started by never using TME+ and then at some point switched to using only TME+. Examining random users, however, reveals several different patterns in their tone usage. Figure 7 shows variations in tone usage for a given emoji by five users, with each user being from one of the five different skin tone groups. These users, and potentially others, do in fact switch between TME+ and TME-, as well as between different tones, even when using the same base emoji.

We characterize this usage as follows. For a particular TME e, a given user u generates a sequence of events E. Each event is a tuple consisting of e and an associated skin tone t_i:

$$E = [(e, t_1), (e, t_2), \ldots, (e, t_n)]$$

Tone modulation occurs when it is the case that $t_{i-1} \neq t_i$, for some fixed emoji e. We can quantify the extent to which tone modulation occurs by examining all E generated by all users and counting instances where $t_{i-1} \neq t_i$. These instances can be divided into two groups: tone-tone (where a user changes tones between TME+), and on-off (where a user switches between a TME+ and TME-, or vice versa).
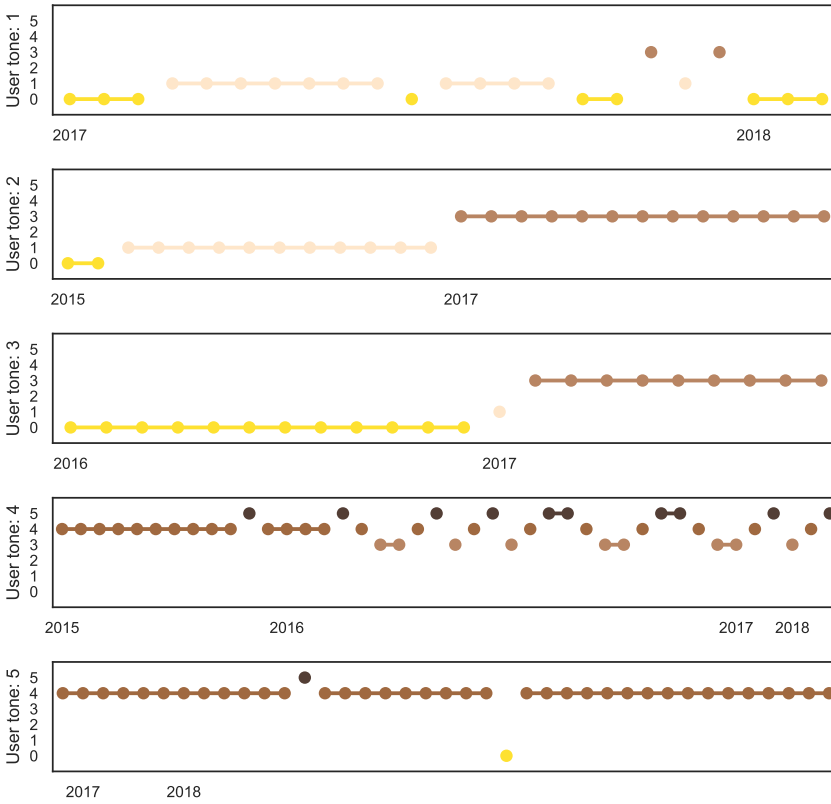
Fig. 7. Examples of tone modulation in production of, from top to bottom, 🤏, 👌, 👏, 🤛 and 🙌 by five different users (selected at random) with skin tones 1 to 5, respectively.

5.1 Quantifying Tone Modulation

To examine the extent of tone modulation in TME users, we present the proportion of tone modulations involving each TME, with and without tones, in Figure 8. These data are aggregated over all four datasets, then divided according to user skin tone group. Each matrix cell shows the proportion of a particular tone modulation in terms of tone before change (rows) and after change (columns). All forms of modulation combined account for 3.89% of events in TME sequences. This makes it a relatively rare phenomenon, supporting the claims of prior work that TME+ are overwhelmingly used for self-representation.

Table 4 aggregates the statistics of each matrix shown in Figure 8, grouping tone modulations into tone-tone (where both t_{i-1} and t_i are TME+: all cells except those in row 0 or column 0), on-off (where t_{i-1} is TME+ and t_i is TME-: all cells in column 0), and off-on (where t_{i-1} is TME- and t_i is TME+: all cells in row 0).

From Figure 8, we see that the single most common event for each user group is an off-on modulation from TME- to the TME+ which most closely matches their own skin tone, as seen in cell $(0, s)$ of Figure 8 where s is the user's actual skin tone. This may be due to users gradually enabling skin tones for individual emoji as they come into use. In general, tone modulation is clustered around the tone associated with the user's actual skin tone and tone-tone changes involving large
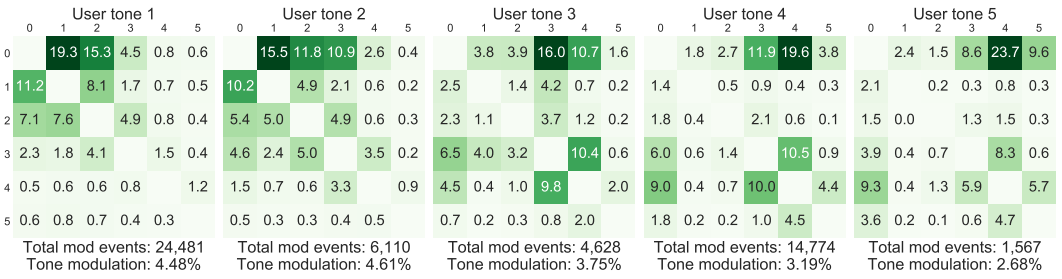
User tone 1

	0	1	2	3	4	5
0		19.3	15.3	4.5	0.8	0.6
1	11.2		8.1	1.7	0.7	0.5
2	7.1	7.6		4.9	0.8	0.4
3	2.3	1.8	4.1		1.5	0.4
4	0.5	0.6	0.6	0.8		1.2
5	0.6	0.8	0.7	0.4	0.3	

Total mod events: 24,481
Tone modulation: 4.48%

User tone 2

	0	1	2	3	4	5
0		15.5	11.8	10.9	2.6	0.4
1	10.2		4.9	2.1	0.6	0.2
2	5.4	5.0		4.9	0.6	0.3
3	4.6	2.4	5.0		3.5	0.2
4	1.5	0.7	0.6	3.3		0.9
5	0.5	0.3	0.3	0.4	0.5	

Total mod events: 6,110
Tone modulation: 4.61%

User tone 3

	0	1	2	3	4	5
0		3.8	3.9	16.0	10.7	1.6
1	2.5		1.4	4.2	0.7	0.2
2	2.3	1.1		3.7	1.2	0.2
3	6.5	4.0	3.2		10.4	0.6
4	4.5	0.4	1.0	9.8		2.0
5	0.7	0.2	0.3	0.8	2.0	

Total mod events: 4,628
Tone modulation: 3.75%

User tone 4

	0	1	2	3	4	5
0		1.8	2.7	11.9	19.6	3.8
1	1.4		0.5	0.9	0.4	0.3
2	1.8	0.4		2.1	0.6	0.1
3	6.0	0.6	1.4		10.5	0.9
4	9.0	0.4	0.7	10.0		4.4
5	1.8	0.2	0.2	1.0	4.5	

Total mod events: 14,774
Tone modulation: 3.19%

User tone 5

	0	1	2	3	4	5
0		2.4	1.5	8.6	23.7	9.6
1	2.1		0.2	0.3	0.8	0.3
2	1.5	0.0		1.3	1.5	0.3
3	3.9	0.4	0.7		8.3	0.6
4	9.3	0.4	1.3	5.9		5.7
5	3.6	0.2	0.1	0.6	4.7	

Total mod events: 1,567
Tone modulation: 2.68%

Fig. 8. Proportion of tone modulations by user skin tone group. Rows indicate a TME's prior tone at t_{i-1}, columns indicate the tone applied when next used at t_i. For example, cell (0,5) shows the proportion of modulations from yellow to the darkest tone. The diagonal (instances of no change) has been removed in order to highlight modulation and is not included in the total number of modulation events used to normalize the proportions in each matrix (listed below as "total mod events"). The tone modulation percentage is the percent of all events that involve tone modulation (i.e., total mod events divided by all events).

Table 4. Summary statistics for forms of tone modulation, based on the data in Figure 8.

User skin tone	Total events	Tone-Tone	On-Off	Off-On
1	24481	9299 (38%)	5292 (21.6%)	9890 (40.4%)
2	6110	2239 (36.6%)	1351 (22.1%)	2520 (41.2%)
3	4628	2198 (47.5%)	765 (16.5%)	1665 (36%)
4	14774	5959 (40.3%)	2951 (20%)	5864 (39.7%)
5	1567	529 (33.8%)	321 (20.5%)	717 (45.8%)

differences are rare. Table 4 shows that overall, tone-tone modulations account for 30-40% of all modulations, with about the same proportion for off-on, and the remaining 20% or so for on-off.

The extent of tone modulation suggests it is not merely accidental. Accidental changes in tone are certainly possible, but given the design of most common interfaces for inputting TME, somewhat unlikely: on iOS, for example, producing 🖐🏿 after having last produced 🖐🏻 involves pressing on 🖐🏻 for a fraction of a second, then dragging to select 🖐🏿 and releasing. It seems unlikely that such an accidental sequence of events can explain all of the observed modulations.

Another explanation considers that users can have multiple devices. They may therefore either have not selected a tone on one device (resulting in on-off events) or have selected a different tone on each device (resulting in tone-tone events), perhaps accidentally or due to the platform rendering differences seen in Figure 4. By constructing E as a sequence of triples (e, t_i, p_i) where p_i represents the platform used for that event (provided by the "source" field as returned by the Twitter API), the number of tone modulations involving two different platforms (where $p_{i-1} \neq p_i$ as well as $t_{i-1} \neq t_i$) is approximately 16,000, around one third of all tone modulations.

Even under the assumption that *all* cross-platform tone modulations are an artefact arising due to the use of multiple devices, and including all of users' one-off tone modulations as accidents, there remain a large number of users generating tone modulation events which are unexplained. The following section investigates these further.

Where is the Unity hiding at???? Smh. 😔✊🏿✊🏼✊🏿✊🏼😢
Don't feel it with conformity remain the exception ✊🏼✊🏿
#BlackWomenMatter ✊🏾✊🏿✊🏿✊🏿

Fig. 9. Examples of "tone rainbows" observed in tweets.

5.2 Characterizing Tone Modulation

We directly examine a subset of tweets and manually classify them based on our overall observations. The aim is to identify factors precipitating tone modulation without making any prior assumptions as to what those might be. We first focus on those tone-tone modulations where users choose a TME+ that is very different from their usual one. Then, we examine on-off modulations where the user produces a TME- for a particular emoji despite having previously, and subsequently, produced it as TME+.

5.2.1 Tone-tone Modulation. Instances of this form of tone modulation can be observed in Figure 7, where emoji vary between multiple possible tones. To investigate these events, we selected all tweets made by the random user group where a user's TME+ differed in tone by at least 2 since the last use of that TME, and the new tone did not match their annotated skin tone—for example, a user annotated with skin tone 2 using 👌🏼 at one time point but 👌🏿 at the next instantiation of that TME. We made this choice on the assumption that a difference of two tones is unlikely to be an accident or random variation between two similar tones that both roughly match the user's own skin tone. The total number of such tone modulations was 1,341, out of 108,584 tweets containing TME+. 48% were made by skin tone (ST) 1 users, 16% by ST2, 26% by ST3, 8% ST4 and 1% ST5.

Tone modulation can occur either *within* or *across* tweets. We found that 15% of tone modulations occur within a single tweet, with 88% made by ST1 users, 1.5% by ST2, 7.8% by ST3, 1.5% by ST4 and 2.4% by ST5. For most of our analysis, we exclude these tweets because we typically found them difficult to interpret. One exception is "tone rainbows"—tweets that apply multiple tones to the same TME in a row. These generally appear to be used to explicitly represent diversity or convey solidarity. The fourth row of Figure 7 comes from a user who has produced such a rainbow for the ✊ emoji, and further examples are illustrated in Figure 9.

Considering only the examples of cross-tweet modulation left 1,136 tweets. Since we are manually inspecting the tweets for their meaning, we then filtered out non-English tweets, leaving 362 tweets. Conservatively assuming that all instances of tone modulation are accidental when there is a platform difference between tweets, our final set for analysis contained 338 tweets from 336 unique users.

We identified five clear categories of tone modulation, plus a sixth category containing miscellaneous usages. A seventh category contains usages which are inscrutable. The proportion of these categories, relative to those tone modulations examined and to the total of TME-containing English tweets from which they were drawn, are shown in Table 5. The categories are now described in detail. Illustrative examples provided from the data are shown in Table 6.

Direct reference: The most common kind of tone modulation involves direct reference to other people. Direct reference involves using the real name of person, including their Twitter username or relevant hashtag in the tweet, or responding to a photo of a person. In all cases, the tone of the TME+ is similar to that of the person or persons being referred to (confirmed for Twitter usernames by looking at the profile photo). The existence of this category is perhaps the most predictable—since TME+ are used for self-representation, the possibility that they can be used for

Table 5. Seven categories of tone modulation distinguished in tweets in the random user group.

Category	Count	% of tone modulations
Direct reference	159	47.0
Oblique reference	61	18.0
Self reference	46	13.6
Group reference	10	3.0
Iconic reference	8	2.4
Miscellaneous	10	3.0
Indeterminate	44	13.0
Total	**338**	**100**

Table 6. Examples of types of reference made using tone modulated TME+. The modulated TME column shows tweet containing a TME+ which is at least two removed from the user's own skin tone. We have replaced names and usernames with placeholders which reflect the skin tone of the referent, where possible.

	User tone	Prior TME usage	Modulated TME
Direct	1	@ST1_user ✌️	@ST5_user @ST4_user Yeah buddy why not!!! 👍
	1	Wow I'm gunna be 25 weeks this week already!! 👶	I have a new niece, $realname 👶💖
Oblique	1	✋ 😊	I hope you get a paper cut ✋🙂
	4	People Always Onna Outside Looking In Like STAY OUT MY MIX ✌️	I can't stop a nigga from doing him & I won't try to either ✌️
Self	1	Go on early runs 🏃 w/ me so I know it's real	@ST2_user Me on pay day 🏧🏃
	3	if your ex still popping up in your notifications or life y'all got unfinished business so stay tf away from me 💁	How somebody feel about me, ain't my business 💁 that shit personal
Group	5	I'm done grabbing leaves cuz 🙋	White girls 🙋 .. #ILoveTheSistas
	4	Just got home. I'm so exhausted 😩😩😩 But very happy that my dad had a safe trip home 🙏	If someone can just get me white privilege for Christmas 🙏
Iconic	1	Chopped off 14 inches of my hair a year ago today & it's all back now 💁	Should I die my hair dark again or go lighter? 💁💁
	1	@ST1_user yep because their fake ass bitches 🙄	Y'all Tuesday is the day 👀💁🙄

representing others is a reasonable extension. What was perhaps less predictable was the differing extents of these two uses: it is extremely uncommon to use toned emoji to refer to other people with skin tones different from that of the user.

Oblique reference: These references are more vague, occurring in tweets which are not in response to other tweets. They contain no real names, usernames or hashtags which could refer to a person. Instead, they characteristically use deictic expressions such as "he" and "they". In some cases these expressions are themselves actually emoji, which is of interest as it suggests another possible source of evidence in support of claims regarding the representative power of emoji. This usage may be related to "vaguebooking", a practice whereby users post deliberately vague or ambiguous messages to social media, and which has been characterized by two seemingly diametrically opposed explanations: users post either to preserve privacy [Child and Starcher 2016] or as a means of gaining attention [Berryman et al. 2017].

Self reference: The referent in these tone modulations appears to be the user, though the TME+ skin tone is at least two removed from the user's tone, i.e. the user has skin tone 1 but the emoji used has skin tone 3. Use of personal pronouns is common and the emoji are often facial, especially 🤷. Unlike direct and oblique reference, there are generally no usernames, real names, hashtags or other indicators of reference to others. Self-reference is further supported by the fact that these emoji can be gendered (in a similar way to how tones are applied) and in all cases emoji gender matches user gender, as determined by user profile photo. It is possible that these usages are input errors, given their rarity, but their co-occurrence with personal pronouns and matching emoji genders suggest this is unlikely to be the case in every instance.

Group reference: These TME+ appear to refer to a group or class of people. The target is non-specific, not targeting any particular person. In our observations, these tone modulations all refer to "whiteness" in some way—white women, white privilege, geographic regions with a lighter-skinned population. The choice of TME+ tone is clearly deliberate and chosen with a specific purpose in mind.

Iconic reference: The visual affordances of some emoji encourage users to make reference to themselves via particular attributes of these emoji. In particular, these are based on differences in hair color (as discussed in Section 4.1) and all involve the user's planned or recent changes to their hair color. The only emoji used in this context are 💇 and 💆. The "hair cut" emoji is predictable, while the "face massage" emoji is likely used here because, on many platforms, it looks like someone having their hair washed by someone else.

Miscellaneous: In this category we include instances of tweets which are song lyrics. In some, but not all, the skin tone more closely matches that of the song's artist. It may be that users have copied and pasted these lyrics, including the emoji, from elsewhere on the internet. A single instance of sexual use was observed, using skin tone 5 to refer to stereotypes.

@user 👳🏿
Wicked wicked wicked 🤟🏻
I'd rather fuck wit myself then waste my time 👌🏼💯
@user happy bday my boy 🤟🏼 much love bro 💯

Fig. 10. Examples of more inscrutable tone modulation

Indeterminate: The remaining instances of tone modulation are somewhat inscrutable, not clearly fitting into the above categories. Examples are shown in Figure 10. In some, reference to another person is made but the tone used in the TME+ matches neither the user nor the referent. These may be input errors, as discussed in Section 5.1. Or, in the case where the message is targeted at another user by including their username, there may be some private meaning to the use of these particular emoji. Since we examined only the preceding tweet rather than any full thread of tweets, it is also possible that some tone modulations in this group could have been classified more accurately into another group if more context were considered.

5.2.2 On-off Modulation. Instances of this form of tone modulation can be observed in rows 1 and 5 in Figure 7, where emoji are realized as TME-, despite having been produced with a tone both immediately before and after. This is the least common form of tone modulation. This is shown by the first column of each matrix in Figure 8 and the summary statistics of Table 4.

To analyze this behavior, we selected all English tweets made by the random user group containing a TME- where the author of the tweet had previously and subsequently used TME+ for that particular

TME. We considered only those instances where the TME+ matched the user's skin tone, in order to select only those instances of on-off modulation which appear to be a conscious decision to turn off tone. Instances where all three tweets were not published via the same platform (e.g. all from the Twitter app, or from the same phone vendor) were removed—approximately 50%.

The total number of on-off tone modulations meeting these requirements was 75, out of 108,584 tweets containing TME+. 48% were made by ST1 users, 22% by ST2, 17% by ST3, 7% ST4 and 6% ST5. Of the 75 tweets, only 60 were amenable to interpretation—the remainder appeared to be from spam accounts with inconsistent TME+ usage.

Of these 60 cases, the majority involve reference to other people. In 18 tweets there is a direct reference by username where the author of the tweet and the person to whom they refer do not have a similar skin tone, based on observation of their profile photos. In 10 tweets, both parties did have a similar skin tone. In 3 tweets which included reference to multiple people, those mentioned had a variety of skin tones. Four tweets made generic reference to people at large and one made oblique reference to a person. There were 6 instances of reference to a Twitter account representing non-humans, such as businesses. In the remaining 14 tweets, which contain no usernames, the reference appears to be solely to the author. We can offer no explanation for this result, though it may be due to using multiple devices of the same type, device reset or input error.

5.3 Discussion

Tone modulation is a rare event, affecting only 4% of the TME observed in 28.3 million tweets. The majority of these cases involve tone-tone modulation, rather than the on-off variety. We restricted our analysis to "extreme" cases of tone modulation, in an effort to identify only the most salient examples of the phenomenon. The major precipitating factor of tone modulation in all cases proved to be reference to other people, either directly by username or real name or indirectly through deixis. There is also evidence that social media users choose to leverage specific visual properties of certain emoji, particularly hair, even when other properties are not congruent with other aspects of their own appearance.

The relatively small number of cases of tone modulation in reference to others, and the fact that many cases of turning off tones entirely appear to be in reference to users with a different skin tone to the tweeter, suggests that some people may be unwilling to use TME+ to refer to others when there is not a common racial/ethnic background. This could be due to the "networked public" [boyd 2007] nature of Twitter. If TME+ are, in Goffman's terminology [Goffman 1959], an important prop for self-expression then to be observed using the wrong prop could be viewed as inauthentic. This would account for the very low levels of tone modulation observed in our data, especially when we consider the replicability issue of social networks like Twitter—users may fear any past examples of such inauthenticity coming back to haunt them in the future. It could even be considered as misappropriation of another group's props and seen as an offensive act, similar to how particular words are restricted to in-group usage.

Misappropriated usage of toned emoji, if indeed offensive, would not show up as negative under the sentiment-based approach used in Robertson et al. [2018], which was purely based on lexical content. Therefore the extent to which negative racially-motivated communication involves TME+ could currently be under-detected. However, the methodology presented here for detecting tone modulation could be adapted to this purpose by relaxing the constraints we placed upon it in order to keep the qualitative analysis tractable.

The rarity of tone modulation should not be interpreted as it being unimportant. This duality, of self and other representation, could be considered in other aspects of emoji understanding. For example, work on interpretation of emoji sentiment shows that not all emoji can be neatly divided into positive, negative and neutral [Novak et al. 2015]. Although such ambiguous emoji
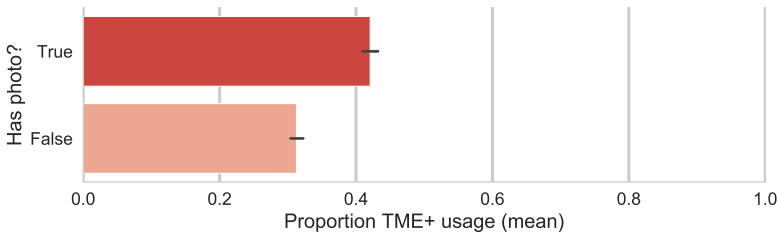
Fig. 11. Average use of TME+ as a proportion of all TME by users with and without profile photos, with error bars showing 95% confidence intervals. Users without a profile photo use TME+ significantly less often.

are not common, they represent a point of inquiry for understanding emoji variation. Such an understanding naturally leads to larger questions, such as the possible linguistic nature of emoji or the influence of culture on emoji interpretation and usage.

6 FURTHER ANALYSES AND GENERAL DISCUSSION

6.1 Are results biased by looking only at users with profile photos?

Our analyses focused on users who publicly display a profile photo. This may be a confounding factor in examining self-representation since these users are, by definition, engaging in self-representation from the outset, which might introduce bias to our findings. Towards addressing this issue we present a comparison of users with and without profile photos, in terms of their TME+ usage.

We divide the random group based on the presence/absence of a valid photo on their profile. Since some user photos classified as invalid in Section 3 may actually show the user, we reclassified those containing multiple people as 'with-photo'. This resulted in 6188 (45.4%) users with a photo and 7430 (54.6%) without. We then computed each user's proportion of TME+ out of all TME used, with the average for each group shown in Figure 11.

An independent samples two-sided t-test confirms a significant difference (t(13616)=11.8, p<0.001) in TME+ usage between users with a profile photo (M=0.42, SD=0.43) and those without a profile photo (M=0.31, SD=0.41). So, users who engage in self-representation through a profile photo also use TME+ more often. More anonymous users, without self-identifying profile photos, do still use TME+ in their online messages, but less frequently.

This difference underscores the concern that other aspects of TME+ usage might be different between the two groups. However, a further analysis suggests that although the no-photo group use TME+ less often, when they do use TME+, they do so in similar ways to the with-photo group. Specifically, for each of the six TME configurations (tones 1-5 or no tone), we computed the proportion of users in each group who had ever used that TME configuration.[5] Results, shown in Figure 12, show that the two groups differ mainly in their overall use of TME- vs TME+ (as shown above). In contrast, the distribution of TME+ usage across the two groups is strikingly similar. Although we cannot directly confirm that the two groups are similar either in terms of the distribution of their skin tones or their choice to use particular tones to represent themselves, taken together these two assumptions do seem to be the simplest explanation for the observed match between the two groups' distribution over TME+ tones. The alternative would require between-group differences in these factors to precisely counterbalance each other in order to produce the observed pattern of results.

---

[5]We did not use counts over all individual TME as this would be dominated by users who produced a lot of TME.
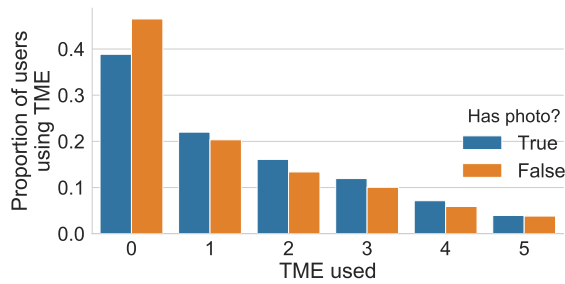
Fig. 12. Proportion of random sample users, with/without a profile photo, who have used one of the six possible TME configurations: zero tone (yellow) or tones 1-5.

This result, then, provides some reassurance that our main claims (that is, emoji tones are important for self-representation, and are used more by darker-skinned users) generalize not only to users from different regions but also to users without photos. Indeed, we would argue that the key factor accounting for both the observed similarities and differences between users with and without photos is self-representation. Displaying a profile photo and using TME+ are both methods of expressing identity online and users who take advantage of the more established form by having a profile photo are also more likely to use the emerging form of expression provided by TME+.

6.2 Implications for emoji designers

The findings of Section 4.2 have ramifications for device vendors, whose emoji designers should account not only for visual affordances [Wiseman and Gould 2018] but also user diversity. The current five-way tone distinction is adapted from the Fitzpatrick photo-typing scale [Fitzpatrick 1988]. This six-way scale classifies skin tone according to how it reacts to UV light, with the emoji modifier version combining the two lightest categories into one. Given the evidence presented here, that the darker end of the emoji tone spectrum is insufficiently granular, how might designers address this issue? A simplistic solution may be to lighten tone 5, but this would merely shift the balance of who suffers from lack of representation. A more inclusive approach would be to add more tone options. From a technical and user interface perspective, this may be challenging to implement. In particular, mobile devices have small screens, which makes it difficult to display a multitude of options at once. However, there have been some interface changes on iOS and other platforms which allow an initial tone selection to propagate to all further applicable situations, or the up-front selection of a tone to be used in all cases. The benefit of overcoming such challenges is clear. It would directly increase the level of representation afforded by emoji—the intended purpose of skin tone modifiers, as stated by the Unicode Consortium [Davis and Edberg 2014].

6.3 Limitations and future directions

We note some limitations with our methodology and analysis. These limitations, necessitated by having to reduce the volume of data for manual inspection, could affect our ability to accurately determine the full scope of tone modulation. In our analysis of tone-tone modulation, we only looked at instances involving maximal changes in tone away from the user's own skin tone. This resulted in a very small sample and prevents us from being able to comment on cases where, for example, ST2 users use tone 1 or 3 in their TME+. Our qualitative results are therefore based on the most extreme examples of tone modulation. As a result, we do not distinguish non-self-reference in cases where the skin tones of the author and the referent are the same. Therefore, any estimates of

the extent of self-referential TME+ can be considered as an upper bound only. Future work should certainly attempt to determine the extent to which users of a given skin tone use TME+ of that same tone to refer to other people. While our results suggest that people avoid referencing another using TME+ tailored to the appearance of the other, it may well be the case that people are more willing to use non-self-referential TME+ when communicating with users with whom they share the same skin tone.

Another practical limitation is restricting our qualitative analysis to English tweets, to aid both our interpretation and that of the reader. Combined with only looking at the random sample, there may be cultural differences in tone modulation which have been obscured. This concern is somewhat alleviated by the results of Section 5.1, which showed there are at least no major differences in tone modulation between users based on their skin tone. Expanding our methodology to locations where skin tone carries especially strong cultural weight, such as South Asia [Baynes 1997], could show the extent to which these cultural views are manifested through emoji usage in an environment where some tones may be considered undesirable.

In addition, we observed many instances of tone modulation which were not readily classifiable as references to other people. These appear to be self-reference but using a skin tone modifier that is unexpected, given the user's actual skin tone. Since we attempted to filter out accidental use by avoiding multi-platform inputs, some of these cases may involve additional precipitating factors which were not detectable given our method for manually inspecting tweets, where we looked only at the preceding and following tweets containing the target TME. These suggest some possible avenues for future work. A more detailed consideration of context might help, but ultimately determining the exact properties of the tone modulation phenomenon will likely require moving toward a more user-focused methodology. Targeted questionnaires and interviews are the way we can hope to answer questions such as why a user chose a particular skin tone in a particular tweet, or why they chose to turn skin tones off for one message in particular. This avenue of investigation is well-motivated by the results presented here, which show that tone modulation is a real aspect of TME usage.

Whether through big data methodology or more targeted interviews, a consideration of audience may be informative. Although Twitter suffers from context collapse as described above, users can nevertheless control the size and nature of their intended audience to some degree by including usernames or hashtags in the tweet. Previous work has shown that Twitter users modulate their use of non-standard lexical items (another way of representing identity) based on audience size and type [Pavalanathan and Eisenstein 2015; Shoemark et al. 2017a,b]. In general, a more restricted audience is likely to be more similar to the user who targets it, due to the homophilic tendencies of social networks like Twitter [Al Zamal et al. 2012; Kwak et al. 2010]. This setting may encourage referring to other people by their appearance, especially if those others are of a similar ethnic background to that of the referrer.

Finally, we note that the data examined here is from a single social media platform and is public in nature. Therefore, the findings here cannot speak to how users behave in private communication, such as in WhatsApp or Messenger groups or in private messages. In Goffman's terminology, this constitutes a different "stage" and therefore may elicit a different performance from people. We can only speculate as to how different these performances may be, if they are indeed distinct from public presentations. Future work could look at WhatsApp groups [Garimella and Tyson 2018] which, while public, are arguably less so than Twitter. However, collecting data in this way raises ethical concerns, due to the blurring of public and private communications.

7 CONCLUSION

Our findings add to the growing body of work on emoji understanding, replicating and extending recent results on the role of emoji skin tone modifiers in self-representation. We set out three main research questions in the introduction. First, do the results of Robertson et al. [2018] generalize to different sub-populations of Twitter users? We showed in Section 4.1 that they do, by extending the findings to users in different regions with different demographic properties (including a majority-Black region). Section 6.1 also provided some evidence that the findings likely generalize to users without profile photos, although these users are less likely to use TME+ overall.

Second, while users' most frequently used emoji tone tends to match their actual skin tone, why is this not always the case? We provide an account in Section 4.2 where we argue that the distribution of the five available tones is not uniform in perceptual space. Compared to the first four tones, the darkest tone is an outlier, and it also obscures visual detail in small emoji. As for skin tones 1 and 2, we point out that these differ in hair color, another aspect of identity that may be more salient than the small difference in skin tone between 1 and 2.

Third, what are the characteristics of emoji use when users select tones which do not reflect their own identity? In Section 5, we presented quantitative and qualitative analyses of two previously unstudied forms of emoji skin tone usage on Twitter—modulation between different emoji skin tones, and deliberately turning emoji tones off altogether. By far the most common situation is for users to choose a tone that matches their own skin tone, suggesting a self-representational use. Nevertheless, by looking at cases where users switched from a self-congruent to incongruent tone, we showed that tones are used not only to represent the self, but also to represent others. We also found preliminary evidence that some users may be reluctant to use tones when referring to others.

Overall, our findings complement prior work which examined emoji only from the perspective of self-representation, and suggest a number of further avenues for exploration to better understand which aspects of themselves and others users choose to represent with emoji, and why they make those choices.

8 ACKNOWLEDGMENTS

A DATA COLLECTION AND ANNOTATION METHODS

Although the Twitter terms of service mean it is not possible to redistribute our data, this appendix describes in detail the method used for annotating data in this paper. By doing so we hope to make it simple for others to replicate and extend our results.

A.1 Selecting users

We used the 1% sample Twitter stream to collect 3.4 million tweets made by 2.6 million unique users on the 14th of March, 2018. From these, users were selected based on the location stated on their profile. We used search criteria that indicated the user was likely to have used Twitter's auto-complete suggestions when personalizing their profile, e.g. showing standard use of capitalization and punctuation.

For Johannesburg, our criteria was simply that this location was listed as "Johannesburg, South Africa". 24,125 matches were found, from which we randomly selected 20,000.

For New York City, the requirement was one of "New York, NY", "Brooklyn, NY", "Manhattan, NY", "Bronx, NY", "Queens, NY" or "Staten Island, NY". 111,746 matches were found, from which we randomly selected 20,000.

For London, we removed all instances of "West", "North", "South", "East" and "Greater" from user locations and then used a regular expression (Figure A.1) to find users in London and its boroughs. 93,466 matches were found, from which we randomly selected 20,000.

```
^London$|^City of London|^City of Westminster|            //
^Kensington and Chelsea|^Hammersmith and Fulham|          //
^Wandsworth|^Lambeth|^Southwark|^Tower Hamlets|           //
^Hackney|^Islington|^Camden|^Brent|^Ealing|               //
^Hounslow|^Richmond upon Thames|^Kingston upon Thames|    //
^Merton|^Sutton|^Croydon|^Bromley|^Lewisham|              //
^Greenwich|^Bexley|^Havering|^Barking and Dagenham|       //
^Redbridge|^Newham|^Waltham Forest|^Haringey|             //
^Enfield|^Barnet|^Harrow|^Hillingdon                      //
```

Fig. 13. Regular expression used for matching user locations to London and its boroughs.

For the random sample, we simply selected 20,000 random users who were not already in one of the other three groups.

A.2 Annotating users

For all 80,000 users we attempted to retrieve their public profile photo. In some cases this was not possible, because users had removed their accounts, set their profiles to private or been banned in the time since initially selecting users. Once we had done this, each group of users was reduced to 10,000 users.

We used the Figure Eight[6] crowdsourcing platform to hire annotators to determine the validity of profile photos and the skin tone of users in valid photos. Users were provided with the task instructions in Figure 14 – these were designed to show examples and valid and invalid photos according to the various criteria. These instructions were available for annotators to view throughout the task. Annotators were paid approximately $0.10 per 40 annotations, with total costs of approximately $700. The platform allows annotators to rate aspects of the task, with this level of pay rated as 2.7 out of 5.

Annotators were shown individual user photos, as in Figure 15, and asked to first determine if the photo was valid and, if so, annotate the photo for skin tone. If the photo was not considered to be valid, the skin tone annotation options would not appear. Skin tone was shown using the "old man" emoji. This emoji was chosen because of the consistent hair color, the lack of distracting visual features and the ease by which the emoji's skin tone could be observed. An option for being unsure about the skin tone was available. Test questions were included to filter out individual annotators who failed easy annotation examples that failed to meet the validity criteria. We did not include any test questions about skin tone, in order to prevent introducing any bias on our part. Annotators who failed to achieve 100% accuracy on these test items were not allowed to progress to the full annotation task. Each photo was annotated by three people.

To select the final Twitter users for our study, we choose only those which had photos marked as valid by two of the three annotators. We then removed any users where any annotator had been

---

[6]http://www.figure-eight.com

Fig. 14. Guidelines on photo validity shown to annotators. Example photos, which were manually selected from random Twitter users not used in this study, are blurred here to protect their anonymity.

unsure of the skin tone in the photo. We then used the Twitter API to select the most recent 3,200 tweets for each user. Users whose profiles were no longer available were removed. We repeated this process again seven months later, once more removing any users whose tweets had since become unavailable to us, even if we had previously been able to collect tweets for them.

Finally, we removed all retweets from the dataset. We performed no additional filtering of users based on language, profile or tweet content.

Fig. 15. The actual annotation interface, with placeholder image. The skin tone question is only visible if the annotator chooses 'yes' for the first question (valid photo).

REFERENCES

Faiyaz Al Zamal, Wendy Liu, and Derek Ruths. Homophily and Latent Attribute Inference: Inferring Latent Attributes of Twitter Users from Neighbors. *ICWSM*, 270:2012, 2012.

Mikiko Ashikari. Cultivating Japanese Whiteness: The 'Whitening' Cosmetics Boom and the Japanese Identity. *Journal of Material Culture*, 10(1):73–91, 2005. doi: 10.1177/1359183505050095. URL https://doi.org/10.1177/1359183505050095.

Francesco Barbieri and Jose Camacho-Collados. How Gender and Skin Tone Modifiers Affect Emoji Semantics in Twitter. In *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics*, pages 101–106, 2018.

Leonard M Baynes. If it's not black and white anymore, why does darkness cast a longer discriminatory shadow than lightness - an investigation and analysis of the color hierarchy. *Denv. UL Rev.*, 75:131, 1997.

Chloe Berryman, Christopher J Ferguson, and Charles Negy. Social media use and mental health among young adults. *Psychiatric quarterly*, pages 1–8, 2017.

Kristina Gavin Bigsby, Jeffrey W Ohlmann, and Kang Zhao. Keeping it 100: Social Media and Self-Presentation in College Football Recruiting. *Big data*, 7(1):3–20, 2019.

Jeff Blagdon. How emoji conquered the world. *The Verge*, March 2013. URL https://www.theverge.com/2013/3/4/3966140/how-emoji-conquered-the-world.

danah boyd. Why youth (heart) social network sites: The role of networked publics in teenage social life. *MacArthur foundation series on digital learning–Youth, identity, and digital media volume*, 119:142, 2007.

Christopher Charles. Skin bleaching in Jamaica: Self-esteem, racial self-esteem, and black identity transactions. *Caribbean Journal of Psychology*, 3(1), 2010.

Andrea Chester and Di Bretherton. Impression management and identity online. *The Oxford handbook of Internet psychology*, pages 223–236, 2007.

Jeffrey T Child and Shawn C Starcher. Fuzzy Facebook privacy boundaries: Exploring mediated lurking, vague-booking, and Facebook privacy management. *Computers in Human Behavior*, 54:483–490, 2016.

Val Colic-Peisker. 'At Least You're the Right Colour': Identity and Social Inclusion of Bosnian Refugees in Australia. *Journal of Ethnic and Migration Studies*, 31(4):615–638, 2005. doi: 10.1080/13691830500109720. URL https://doi.org/10.1080/13691830500109720.

Mark Davis and Peter Edberg. Proposed Draft Unicode Technical Report: Unicode Emoji. Technical Report 51, Unicode Consortium, December 2014. URL https://www.unicode.org/reports/tr51/tr51-1-archive.html.

M Dickey. Thoughts on white people using dark-skinned emoji. *TechCrunch*, October 2017. URL https://techcrunch.com/2017/10/01/thoughts-on-white-people-using-dark-skinned-emoji/.

TB Fitzpatrick. The validity and practicality of sun-reactive skin types I through VI. *Archives of Dermatology*, 124(6):869–871, 1988. doi: 10.1001/archderm.1988.01670060015008. URL +http://dx.doi.org/10.1001/archderm.1988.01670060015008.

Kiran Garimella and Gareth Tyson. Whatapp Doc? A First Look at Whatsapp Public Group Data. In *Twelfth International AAAI Conference on Web and Social Media*, 2018.

Erving Goffman. The presentation of self in everyday life. *New York*, 1959.

Sherri Grasmuck, Jason Martin, and Shanyang Zhao. Ethno-Racial Identity Displays on Facebook. *Journal of Computer-Mediated Communication*, 15(1):158–188, 10 2009. ISSN 1083-6101. doi: 10.1111/j.1083-6101.2009.01498.x. URL https://doi.org/10.1111/j.1083-6101.2009.01498.x.

Bernie Hogan. The presentation of self in the age of social media: Distinguishing performances and exhibitions online. *Bulletin of Science, Technology & Society*, 30(6):377–386, 2010.

Tianran Hu, Han Guo, Hao Sun, Thuy-vy Thi Nguyen, and Jiebo Luo. Spice up Your Chat: The Intentions and Sentiment Effects of Using Emoji. *CoRR*, abs/1703.02860, 2017. URL http://arxiv.org/abs/1703.02860.

George H. Joblove and Donald Greenberg. Color Spaces for Computer Graphics. In *Proceedings of the 5th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '78, pages 20–25, New York, NY, USA, 1978. ACM. doi: 10.1145/800248.807362. URL http://doi.acm.org/10.1145/800248.807362.

Sanja Kapidzic and Susan C Herring. Race, gender, and self-presentation in teen profile photographs. *New Media & Society*, 17(6):958–976, 2015.

Linda K Kaye, Helen J Wall, and Stephanie A Malone. "Turn that frown upside-down": A contextual account of emoticon usage on different virtual platforms. *Computers in Human Behavior*, 60:463–467, 2016.

Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is Twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, pages 591–600. AcM, 2010.

Linda Kathryn Larkey, Michael L. Hecht, and Judith Martin. What's in a Name?: African American Ethnic Identity Terms and Self-Determination. *Journal of Language and Social Psychology*, 12(4):302–317, 1993. doi: 10.1177/0261927X93124003. URL https://doi.org/10.1177/0261927X93124003.

Katie Lebel and Karen Danylchuk. How tweet it is: A gendered analysis of professional tennis players' self-presentation on Twitter. *International Journal of Sport Communication*, 5(4):461–480, 2012.

Hannah Miller, Daniel Kluver, Jacob Thebault-Spieker, Loren Terveen, and Brent Hecht. *Understanding emoji ambiguity in context: The role of text in emoji-related miscommunication*, pages 152–161. AAAI Press, 2017.

Petra Kralj Novak, Jasmina Smailović, Borut Sluban, and Igor Mozetič. Sentiment of emojis. *PloS one*, 10(12):e0144296, 2015.

Office for National Statistics. *2011 Census aggregate data*. 2011. doi: http://dx.doi.org/10.5257/census/aggregate-2011-1.

Zizi Papacharissi. Without you, I'm nothing: Performances of the self on Twitter. *International journal of communication*, 6: 18, 2012.

Umashanthi Pavalanathan and Jacob Eisenstein. Audience-modulated variation in online social media. *American Speech*, 90 (2):187–213, 2015.

Jean S. Phinney and Linda Line Alipuria. "Ethnic identity in college students from four ethnic groups". *Journal of Adolescence*, 13(2):171 – 183, 1990. ISSN 0140-1971. doi: https://doi.org/10.1016/0140-1971(90)90006-S. URL http://www.sciencedirect.com/science/article/pii/014019719090006S.

Jean S Phinney and Anthony D Ong. Conceptualization and measurement of ethnic identity: Current status and future directions. *Journal of counseling Psychology*, 54(3):271, 2007.

Jessica Ringrose. Are you sexy, flirty, or a slut? Exploring 'sexualization'and how teen girls perform/negotiate digital sexual identity on social networking sites. In *New Femininities*, pages 99–116. Springer, 2011.

Alexander Robertson, Walid Magdy, and Sharon Goldwater. Self-Representation on Twitter Using Emoji Skin Color Modifiers. In *Proceedings of the 12th International Conference on Web and Social Media, ICWSM 2018*. AAAI Press, 2018.

Sashank Santhanam, Vidhushini Srinivasan, Shaina Glass, and Samira Shaikh. I Stand With You: Using Emojis to Study Solidarity in Crisis Events. In *Proceedings of the First International Workshop on Emoji Understanding and Applications in Social Media*, 2018.

Bryan Semaan, Bryan Dosono, and Lauren M Britton. Impression Management in High Context Societies:'Saving Face'with ICT. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 712–725. ACM, 2017.

Philippa Shoemark, James Kirby, and Sharon Goldwater. Topic and audience effects on distinctively Scottish vocabulary usage in Twitter data. In *Proceedings of the Workshop on Stylistic Variation at EMNLP 2017*, Copenhagen, Denmark, 2017a.

Philippa Shoemark, Debnil Sur, Luke Shrimpton, Iain Murray, and Sharon Goldwater. Aye or naw, whit dae ye hink? Scottish independence and linguistic identity on social media. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, volume 1, pages 1239–1248, 2017b.

Statistics South Africa. Census 2011, 2011. URL http://www.statssa.gov.za/?page_id=3836.

Garreth W. Tigwell and David R. Flatla. Oh That's What You Meant!: Reducing Emoji Misunderstanding. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, MobileHCI '16, pages 859–866, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-4413-5. doi: 10.1145/2957265.2961844. URL http://doi.acm.org/10.1145/2957265.2961844.

United States Census Bureau. Census 2010, 2010. URL https://www.census.gov/quickfacts/fact/table/newyorkcitynewyork/PST045217.

Michael Wesch. Youtube and you: Experiences of self-awareness in the context collapse of the recording webcam. *Explorations in media ecology*, 8(2):19–34, 2009.

Sarah Wiseman and Sandy J. J. Gould. Repurposing Emoji for Personalised Communication: Why 🥀 Means "I Love You". In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 152:1–152:10, New York, NY, USA, 2018. ACM. ISBN 978-1-4503-5620-6. doi: 10.1145/3173574.3173726. URL http://doi.acm.org/10.1145/3173574.3173726.

Megan A Wood, William M Bukowski, and Eric Lis. The digital self: how social media serves as a setting that shapes youth's emotional experiences. *Adolescent Research Review*, 1(2):163–173, 2016.

Shanyang Zhao, Sherri Grasmuck, and Jason Martin. Identity construction on Facebook: Digital empowerment in anchored relationships. *Computers in human behavior*, 24(5):1816–1836, 2008.

J Zimmerman. Racially diverse emoji are a nice idea. But will anyone use them? *The Guardian*, April 2015. URL https://www.theguardian.com/commentisfree/2015/apr/22/racially-diverse-emoji-are-a-nice-idea-but-will-anyone-use-them.