Identifying the Risks of Miscommunicating with Emoji

A THESIS
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

Hannah Miller

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Loren Terveen and Brent Hecht

August 2018

Acknowledgements

Despite my use of the pronoun "I" in this thesis, the research below would not have been possible without the support and collaboration of many colleagues. The people who have contributed to the research in this thesis include:

Loren Terveen, Brent Hecht,
Jacob Thebault-Spieker, Shuo Chang, Isaac Johnson,
Daniel Kluver, and Zachary Levonian

I am so grateful to my advisors, Loren Terveen and Brent Hecht, for their support and guidance. I have learned and grown so much under their advisement.

On December 4, 2015, I decided to go to a lab research meeting that I did not typically attend. I was engaged with the research topic that was being pitched, and I am so fortunate that I was elected to lead the project. Thank you to Jacob Thebault-Spieker, Shuo Chang and Isaac Johnson for not only trusting me with the project, but also supporting me with it along the way. I learned so much from each of you and will always cherish that collaborative experience. Since then, I have also been fortunate to work with Daniel Kluver and Zachary Levonian. Altogether, this work would not be what it is without this excellent group of colleagues.

The work in this thesis took place in GroupLens Lab, of which I am beyond grateful to have been a member. Thank you to GroupLens for being such a strong source of collaboration, support and friendship over the past few years, and in the future.

Finally, I would also like to specifically thank my committee: Loren Terveen, Brent Hecht, Haiyi Zhu and Yuqing Ren. It has been a pleasure presenting and discussing this work with these individuals.

Dedication

This thesis is dedicated to Bart Hillberg and my family, Lynn, Jake and Haley Miller (and Joe). Bart, Mom, Dad and Haley, without your support over the years, none of the pages below would have been written.

Abstract

Emoji have become commonplace in nearly all forms of text-based computer-mediated communication, but as picture characters with nuanced details, emoji may be open to interpretation. Emoji also render differently on different viewing platforms (e.g., Apple's iPhone vs Google's Nexus phone), potentially leading to communication errors. It is unknown whether people are aware that emoji have multiple renderings, or whether they would change their emoji-bearing messages if they could see how these messages render on recipients' devices. In this thesis, I identify the risks of miscommunicating with emoji. Drawing from psycholinguistic theory, my collaborators and I developed a measure to demonstrate the potential for misconstrual of emoji due to people varying in their interpretations. I also investigated whether the presence of text would reduce this potential, finding little to no support for this hypothesis. Finally, I explored the real-world impact of the multi-rendering nature of emoji, finding that a substantial proportion of people are unaware that emoji have multiple renderings and that, in many instances of emoji use, increased visibility of different emoji renderings would affect communication decisions. To provide this visibility, I developed emoji rendering software that simulates how a given emoji-bearing text renders on various platforms, including when platforms do not support the given emoji. Altogether, this work identifies the risks of miscommunicating with emoji, but it also informs the design and development of technology to, at least partially, mitigate these problems. The data I produced and the emoji rendering software I built can be integrated into new tools for communication applications to prevent regretful exchanges due to ambiguous emoji or emoji rendering differences across platforms.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Emoji are "picture characters" (literal translation from Japanese) that have become commonplace in nearly all forms of text-based computer-mediated communication, including smartphone texting [41], social media sharing [19], and advertising [68]. Hundreds of millions of people interact with emoji on a daily basis, whether as authors, recipients, or both. As an indicator of the ubiquity of emoji, Oxford Dictionaries declared the "face with tears of joy" emoji (😂)[1] to be the 2015 "word of the year" [73], noting that "emoji have come to embody a core aspect of living in a digital world that is visually driven, emotionally expressive, and obsessively immediate" [73].

The Unicode Consortium provides a worldwide text-encoding standard for emoji characters just as it does for more traditional characters (e.g., Roman alphabet letters, numbers, Chinese characters) [74]). The Unicode standard provides a *code point* (or sequence of code points) and a name for each emoji character, but it is unlikely that people recognize emoji characters by these identifiers (i.e., 😁 is not usually described as the "Beaming Face with Smiling Eyes" emoji or "U+1F601"). Rather, as a picture character, an emoji conveys its meaning through its graphic resemblance to a physical object (e.g., a smiling face). But it is not well understood how people interpret the

---

[1] The emoji renderings included in the text are Apple's renderings, unless otherwise specified.

meaning of emoji. Words have a dictionary definition, but emoji are nuanced, visually-detailed graphics that may be more open to interpretation.

Furthermore, graphics for emoji characters are *not* standardized by the Unicode Consortium. Instead, the appearance of an emoji character is *rendered* by a font. Critically, emoji fonts are largely specific to individual technological *vendors.* This means that emoji look different on devices or applications from different vendors (e.g., Apple, Google; see Figure 1.1a). In other words, when communicating with emoji, a receiver will see different emoji renderings than the sender if they are using devices from different vendors. Emojipedia, an emoji reference website, currently tracks 12 vendors that each have their own emoji fonts [75].

Vendor emoji differences, however, only describe one part of the complexity of the emoji rendering ecosystem. Additional complexity is added by the fact that vendors update their emoji fonts over time, along with their other operating system or application updates. As such, emoji fonts are actually *vendor-version* specific, not just vendor-specific (see Figure 1.1b). For example, a sender with an Android phone using version 8.1 would see a different rendering of the emoji in Figure 1.1b than a recipient with an Android phone using version 7.0, even though both of these devices use an operating system from Google. However, a recipient with an Android phone using version 8.0 would see the same rendering as the sender, because Google did not update the emoji character in Figure 1.1b in its Android update from 8.0 to 8.1. To clarify, sometimes

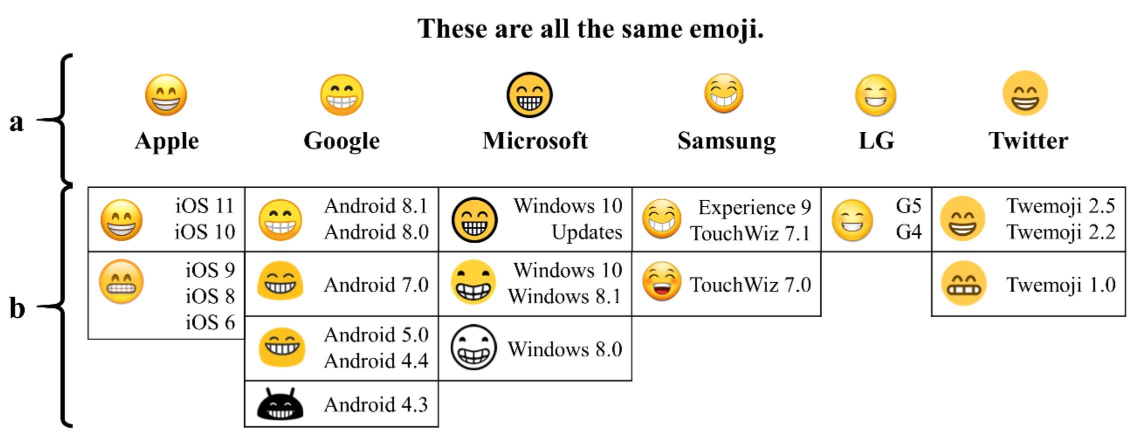


Figure 1.1: (a) Vendor-specific renderings of the "Beaming Face with Smiling Eyes" emoji (U+1F601) **(b)** Vendor-version specific renderings of the same emoji, in reverse chronological vertical order.

vendors do not update every emoji character in a given emoji font update (e.g., see cells in Figure 1.1b with more than one version listed), so the same emoji rendering may belong to more than one version's emoji font. In my work, I use the term *platform* to refer to a device or application using a specific vendor-version configuration (i.e., emoji font), and an emoji *rendering* is the graphic of an emoji character from the emoji font of the platform used to produce it. Emojipedia currently tracks over 50 vendor-version configurations (i.e., emoji fonts), which means any given emoji character may have 50 different renderings (though some may no longer be in use, e.g., if every person has updated their device(s) from a given vendor past a given version[2]).

Given the platform-dependence of emoji, communicating with emoji can take place either *within platform* or *across platforms*. An author communicating with an emoji selects and shares it via a mobile or desktop platform. Likewise, a recipient views the emoji via a mobile or desktop platform. If the author and recipient are using the same platform, then they are communicating *within platform* and they see the same emoji rendering. If the author and recipient are using different platforms, then they are communicating *across platforms* and see different renderings of emoji. Though I make this distinction between within- and cross-platform contexts of emoji use, people only see the emoji renderings on the mobile or desktop platform they are using. This means people may not be aware that it is possible they are seeing different renderings than their communication partners, let alone that emoji have multiple renderings.

Since emoji are a form of digital text for communication, I contextualize emoji use in Herbert Clark's psycholinguistic theory of language use [12]. In psycholinguistics, a *construal* is the way that an individual interprets communication. That is, when a speaker communicates something, the addressee interprets or *construes* what s/he believes the speaker to mean. When the addressee's interpretation differs from what the speaker intended, a *misconstrual* occurs. Thus, in order for emoji to be properly construed, people need to interpret emoji renderings the same way (within platform), even though, as picture characters, they may be more open to interpretation. Also, because emoji have

[2] This adds yet another dimension of complexity, because it is very difficult to determine which vendor-version configurations are in use (including new versions getting released).

multiple renderings, people need to interpret the different renderings of the same emoji character the same way (across platforms). Further, they have to do this without seeing the other renderings, if they even know that other renderings exist.

Motivated by psycholinguistic theory, my thesis focuses on identifying the risks of miscommunicating with emoji. Drawing from psycholinguistic theory, my collaborators and I developed a measure to demonstrate the potential for misconstrual of emoji due to people varying in their interpretations. I also investigated whether the presence of text would reduce this potential, finding little to no support for this hypothesis. Finally, I explored the real-world impact of the multi-rendering nature of emoji, finding that a substantial proportion of people are unaware that emoji have multiple renderings and that, in many instances of emoji use, increased visibility of different emoji renderings would affect communication decisions.

More specifically, in the first chapter of my thesis work, I explored whether emoji renderings and/or their differences across platforms give rise to diverse interpretations of emoji. Through an online survey, I solicited 304 people's interpretations of a sample of the most popular emoji characters, each rendered for multiple platforms. I calculated the variance in interpretation of the emoji in terms of both sentiment and semantics, quantifying which emoji are most (and least) likely to be misinterpreted. From this analysis, I concluded that using emoji carries a substantial risk for misconstrued communication, both within and across platforms [43].

While my first project examined variation in the interpretation of emoji, this focused on standalone emoji, meaning the emoji were interpreted in isolation. Although emoji sometimes are used in isolation, most often they are accompanied by surrounding text [41]. I hypothesized that examining emoji in textual context would reduce the observed potential for miscommunication, so I conducted another similar survey of 2,482 people's interpretations of emoji, now also considering emoji in textual contexts (tweets). I ultimately found little to no support for my hypothesis. Instead, my work showed that in general, emoji are not significantly less ambiguous when interpreted in context than when interpreted in isolation. In other words, supplementing emoji usage with text does not necessarily lower the risk of miscommunication [42].

The final chapter of my thesis provides evidence that surfacing the multi-rendering nature of emoji would have meaningful effects on real-world text communication. Though I found in my prior work that people vary in their interpretations of different emoji renderings, it is not known whether people are even aware that emoji have multiple renderings, since people can only see the emoji renderings specific to the platform they are currently using. More critically, it is unclear if people would change their communication behavior if they could see how their messages rendered on other platforms. To address these open questions, I needed a way to show emoji rendering differences across platforms to people in the context of their own messages, and a tool to do this did not exist. As such, I developed emoji rendering software that parses emoji from input text and accurately displays how the text would render on a wide variety of platforms.

I embedded this software in an online survey deployed on Twitter so that I could use participants' own emoji-bearing tweets to expose the multi-rendering nature of emoji. At least 25% of my 710 survey respondents were not aware that emoji have multiple renderings. This suggests that a substantial proportion of people do not know that the emoji renderings they see are not always the same as the renderings their communication partners see. Additionally, 20% of my respondents indicated that they would have edited or not sent their emoji-bearing tweet if they had known how the tweet rendered on different platforms. Generalizing to the tweet population, this means I estimate that millions of such potentially regretful tweets are shared per day, because people currently are not afforded visibility of emoji rendering differences across platforms.

My thesis motivates the need for new technology to better support people as they communicate with emoji. This need is exacerbated by the fact that intellectual property concerns and branding incentives will likely ensure that emoji rendering differences across platforms will persist in the foreseeable future [24]. I propose building tools that provide an emoji "preview" function similar to that in my final study. Such tools would give people the awareness and visibility they currently lack while communicating with emoji.

In summary, my thesis work makes the following contributions:

- I identified potential for miscommunication of standalone emoji, both in terms of sentiment and semantics and both within and across platforms.
- I compared the potential for sentiment miscommunication of emoji in isolation versus in natural textual contexts, finding little to no support that emoji are less ambiguous in context.
- I produced the first empirical information on the general awareness of the multi-rendering nature of emoji, observing that at least 25% of the Twitter users I surveyed were not aware.
- I developed emoji rendering simulation software that affords visibility of emoji rendering differences across platforms in the context of a given text, including when platforms do not support the given emoji.
- I quantified the proportion of emoji-bearing tweets whose authors would prefer to not send as-is after seeing the tweet rendered across platforms, which allowed me to estimate the real-world effect of people not being able to see emoji rendering differences across platforms.

Altogether this work identifies the risks of miscommunicating with emoji. Below, the work of this thesis is situated in related work in the field. The following three chapters describe the work in detail. The final chapter concludes with a discussion of implications for design and future work dedicated to reducing the risk of miscommunicating with emoji and helping people navigate the multi-rendering emoji ecosystem.

Chapter 2

Related Work

2.1 Emoticons, the Predecessor to Emoji

Emoticons are "typographic symbols that appear sideways as resembling facial expressions," [63] such as :). They have been in use in text-based communication since at least the early 1980s, with numerous studies documenting their prevalence in SMS texts [62], blogs [31], and, more recently, Twitter [47].

Much research has focused on the role that emoticons can play in complementing traditional text-based computer-mediated communication (CMC). Derks et al. concluded in a survey of emotion in CMC that the function of emoticons in digital text largely parallels non-verbal cues in face-to-face communication [17]. With respect to interpretation, Walther and D'Addario found that while the emotional valence of text (e.g., "I am happy") tends to be more important than any accompanying emoticons, a negative emoticon (e.g., :("frowny face") can significantly change the interpretation of the message [63]. Lo provided additional evidence that emoticons affect interpretation, showing that the same text can be perceived as either happy or sad depending on which emoticon accompanies it [37]. Going beyond interpretation of individual messages, Liebman and Gergle demonstrated that emoticons (along with punctuation) are important in interpersonal relationship development over text-based communication [35]. Together, this work emphasizes that emoticons play an important role in text-based communication, affecting interpretation and interpersonal relationships.

2.2 The Rise of Emoji

Emoji were first created in the late 1990s in Japan but were not officially added to the Unicode Standard until 2009 [16]. They have become quite popular since then, with, for example, nearly half of text on Instagram [19] containing emoji. As another example, over 60 million emoji are sent every day on Facebook, with an average of 5 billion sent every day on Messenger [53]. I observed in my research that about 16 percent of all tweets contain emoji, so about 80 million tweets with emoji are shared daily.

Emoji are often described as a successor to emoticons (e.g., [45]), and Pavalanathan and Eisenstein found that while emoticons are decreasing in popularity on Twitter, emoji are increasing in popularity and seem to be replacing, not complementing, emoticons [48]. Early emoji research indicates that emoji do fulfill much the same role of complementing digital text [14]. Kelly and Watts interviewed a culturally diverse group of people and found that they did use emoji in text-based communication to convey and modify the meaning and emotional valence of their words [33].

Interest in emoji in the computing research community has increased dramatically in the past few years. Computing researchers have been focusing on topics ranging from functions of emoji [1,14,28,33,44] to emoji usage patterns [2,11,38] to applications for sentiment classification and text understanding [3,21,25,45]. The work in this thesis, however, is the first to directly study the potential for miscommunication associated with emoji. The sections below detail the strains of emoji-related literature most relevant to the research in this thesis.

2.3 Consistency of Emoticon and Emoji Interpretation

Whereas the display of emoji is platform vendor-dependent, emoticons, as text, are displayed relatively consistently. Walther and D’Addario found high agreement across their participants (226 mostly male students) around sentiment interpretations of the three emoticons that they studied, :-) and :-(and ;-) [63]. In research on using emoticons in sentiment analysis, Davidov et al. found that when Amazon Mechanical Turk participants were presented with tweets in which emoticons had been removed, they were able to identify with high precision the original emoticon that had been in the tweet [15].

Less is known about the consistency of emoji interpretation. Researchers such as Liu et al. [36], Novak et al. [45], and Kimura and Katsurai [34] have developed classifiers of emoji sentiment by labelling emoji with the sentiment of the surrounding text. These projects found instances of emoji being associated with different, and occasionally opposite, sentiment labels. Building on this work and the work in this thesis, Wijeratne et al. provide a similar resource *EmojiNet* [76] but for semantics [65,66]. This "sense inventory" also associates multiple meanings with individual emoji [65,66]. These efforts show that emoji may be used in different ways and take on different meanings, but they do not address whether people agree on the meaning of an emoji in a given use case. The research in this thesis addresses this gap in the literature.

Importantly, the resources from the above efforts do not differentiate emoji renderings, only emoji characters. In other words, the sentiment classifiers and sense inventories that have been built for emoji associate meaning with each emoji character, not each emoji rendering. After my first study, Wijeratne et al. [66] acknowledged this crucial facet of emoji, mentioning that the senses identified in their inventory for a given emoji may be more or less associated with its different renderings. In fact, they explored this hypothesis for a subset of 40 emoji, ultimately finding that this was true for the majority [66]. However, this finding is not reflected in the sense inventory since the inventory does not differentiate emoji renderings, only emoji characters. The research in this thesis was the first to explicitly consider that emoji have multiple renderings and to study how this characteristic of emoji might further complicate interpretation.

Despite the progress made by the above literature, no work prior to the work in this thesis investigated how the interpretation of emoji varies, nor how emoji rendering differences across platforms might contribute to this variance. It also remains unknown whether people are even aware that emoji have multiple renderings, and for those that are aware, whether they perceive enough difference between renderings to change their messaging behavior. This thesis addresses these gaps in the literature.

It is important to note that the critical importance of understanding user behavior around emoji rendering differences across platforms is bolstered by legal and economic factors that make it highly unlikely that cross-platform communication will disappear anytime in the foreseeable future [24]. Specifically, vendors are incentivized to create a distinct style for their emoji renderings as a way to build brand loyalty [24], as well as to take advantage of opportunities to incorporate their brand into their emoji renderings [6]. Also, emoji renderings may be protected by intellectual property rights, including copyright, trademark, design patents, and publicity rights [24]. These factors prevent vendors from using or adopting each other's emoji renderings, thereby preventing complete convergence of emoji fonts. In fact, recent events provide evidence of this prevention [7]: Slack, a popular communication platform for group collaboration, recently switched from rendering emoji using a single emoji font (Apple's) to rendering emoji natively (i.e., using the viewing platform's emoji font) [71]. That is, Slack went from being a within-platform communication setting (using a single emoji font) to a cross-platform setting, and this was likely due to the legal and economic factors mentioned above. Such factors reduce possible emoji convergence efforts to individual vendors, e.g., improving version updates (e.g., getting people to update their devices and/or applications) or creating cross-platform applications that only use the vendor's emoji font (e.g., Facebook Messenger [77]), that is, of course, in the rare case that the vendor has its own emoji font (unlike Slack [71]).

2.4 Technology to Support Emoji Use

Most technology (and informing research) to support people as they use emoji serve purposes other than managing the potential for emoji-related miscommunication. Such tools and efforts include various designs for emoji selection [49,50] (e.g., various emoji keyboards [78–82]) and search [22,83], emoji prediction based on what is typed [57,67,70,72,84,85], and emoji "translation" to convert text to emoji [5,85–87]. The sentiment lexicons [34,45] and semantic models [76] representing emoji meaning described in the previous section are relevant to the potential for emoji-related miscommunication, because emoji may be associated with a range of meaning. However, these resources are intended to assist machines in automatic text processing rather than to assist people while they are communicating.

The best resources that currently exist for people to consult regarding the multiple renderings of emoji are websites that maintain repositories of renderings associated with each emoji character for some set of vendors. Many such websites, including the full official Unicode emoji list, do not maintain historical versions of renderings, even though many such older versions are still in use. Likewise, many such websites are outdated, given that individual vendors act independently and update relatively frequently, in some cases every few months (e.g., Microsoft updated emoji in August 2016, April 2017, October 2017, and April 2018 [88]). To my knowledge, Emojipedia.org [75] is the most comprehensive inventory of emoji, maintaining information from most platform vendors as well as most historical versions. However, if one is using Emojipedia to look up what a given communication will look like across platforms, one can only do so out of context. Additionally, doing this look-up for each emoji in each message is an excessive burden given the number of emoji-bearing messages that are sent by people each day.

Overall, the lack of technology to support people as they communicate with emoji means that they are almost always “flying blind” when it comes to managing the multi-rendering nature of emoji, if they even know to manage it at all.

Chapter 3

Varying Interpretations of Emoji

An emoji conveys its meaning through its graphic resemblance to a physical object (e.g., a smiling face), but it is not well understood how people interpret the meaning of emoji. Words have a dictionary definition, but emoji are nuanced, visually-detailed graphics that may be more open to interpretation. Furthermore, since emoji render differently on different platforms, the emoji graphic that is sent by one person on one device may be quite different than what is seen by the recipient using a different device.

In this chapter of my thesis, I worked to identify the risk of miscommunicating with emoji. Recall from the introduction that when an addressee's interpretation differs from what the speaker intended by a given utterance, a *misconstrual* occurs [12]. Thus, to study the risk of miscommunicating with emoji, I explored and quantified their potential for misconstrual. Given an emoji utterance, misconstrual can arise from differing interpretations derived from either (1) the same rendering, in a within-platform communication context or (2) different renderings, in a cross-platform communication context. As such, I broke down my study of the potential for misconstrual of emoji into two research questions based on within- and cross-platform communication contexts:

> **RQ1 (Within Platform)**: Do people look at the exact same rendering of a given emoji and interpret it the same way? For each platform, which emoji are most/least likely to be misinterpreted in communication within platform?

> **RQ2 (Across Platforms)**: Do people interpret one platform's rendering of an emoji character the same way that they interpret a different platform's rendering? Which emoji are most/least likely to be misinterpreted in communication across platforms?

Using an online survey, I solicited people's interpretations of a sample of the most popular anthropomorphic (i.e., human-looking) emoji characters. In order to analyze how emoji interpretations vary for renderings across platforms, the survey included renderings of each emoji from five major mobile vendors: Apple, Google, Microsoft, Samsung, and LG. In the survey, people interpreted a sample of emoji renderings by judging the sentiment (i.e., how positive is this emoji?) and providing open-ended semantic responses (i.e., what does this emoji mean?) for each. Multiple participants (median 37) interpreted each emoji rendering so that I could observe the variation among these interpretations.

To analyze this data, I needed a metric to quantify the potential for misconstrual associated with each emoji rendering, as well as with each emoji character across renderings. The distance between two participants' interpretations of a given emoji rendering captures the degree to which these two participants disagree on the meaning of that rendering. If these two participants were communicating with each other, this would quantify the *misconstrual* or miscommunication that occurs when exchanging this emoji rendering. Therefore, by computing the distances between all interpretations of a given emoji rendering, the average of these distances represents the average misconstrual of the rendering. The higher this average *misconstrual score*, the more potential this emoji rendering has to be misconstrued, and thus the more risk it poses to communication. To consider communicating with emoji across platforms, I performed the same computation except with pairs of participant interpretations from two different renderings of the same emoji character.

I found that only 4.5% of emoji symbols I examined have consistently low variance in their sentiment interpretations. Conversely, in 25% of the cases where participants rated the same rendering, they did not agree on whether the sentiment was positive, neutral, or negative. When considering renderings across platforms, these disagreements only increased. For the "grinning face with smiling eyes" emoji (U+1F601), participants described the Google rendering 😁 as “blissfully happy” while the Apple rendering 😁 was described as “ready to fight.” This divergence was reflected in this emoji’s sentiment results: on average people interpret Google’s rendering to be positive, Apple’s to be negative. I conclude that emoji usage may be ripe for misconstrued communication and provide implications for design to manage the likelihood of misinterpretation when using emoji.

3.1 Survey Study

I created an online survey to solicit people’s interpretations of a sample of emoji characters, each rendered for multiple platforms. This section details the emoji and platform selection, as well as the survey design, participants and the data collected for analysis.

3.1.1 Emoji Character Sample

I selected a sample of the most popular emoji characters. To determine their popularity, I identified emoji present in a dataset of approximately 100 million random tweets collected between August and September 2015. This dataset provided a recent ranking of how often each emoji is used.

I restricted my sampling to anthropomorphic emoji, or those that represent faces or people, because (1) they are very common and (2) I hypothesized that misconstrual would be more likely among these emoji than those that characterize “things” (e.g., an airplane, a balloon, flowers, flags, etc.). Anthropomorphic emoji account for approximately 50% of emoji use in the Twitter dataset, and SwiftKey reports that faces or smileys comprise 59% of emoji characters typed with their smartphone keyboard app [58]. I selected the top 25 most popular anthropomorphic emoji characters for my sample.

3.1.2 Platform Selection

To investigate how people interpret renderings from different platforms, I solicited people’s interpretations of multiple platform renderings of each emoji character in my sample, focusing on smartphone platforms. Using comScore reports from 2015 [69], I picked the top three smartphone platform vendors: Android, Apple, and Microsoft. Since Android is fragmented by manufacturer, I selected Google's rendering, as well as the renderings of the top two Android hardware manufacturers: Samsung and LG.[3] I used the current renderings[4] from these five vendors for every emoji character in my study. To collect the graphics of the emoji to use in my survey, I used Emojipedia [75].

3.1.3 Survey Design

With 5 platform renderings of 25 emoji characters, I gathered survey results for 125 total emoji renderings. I employed a purely random between-subjects design, and each participant received a random sample of 15 emoji renderings to interpret from the 125 total. I aimed to collect approximately 40 interpretations per emoji rendering. Thus for a total of 5000 interpretations, and 15 interpretations per participant, I recruited 334 participants to complete the survey.

The survey began with a section to solicit background information about the participants such as their age, their gender, the smartphone platform that they use, and their frequency of emoji usage. Next, each emoji rendering was displayed on its own survey page, which showed an image of the emoji and asked:

1. In 10 words or less, say what you think this emoji means:
2. If you had to use one or two words to describe this emoji, which would you use?
3. Judge the sentiment expressed by the emoji [on an ordinal scale from Strongly Negative (-5) to Strongly Positive (5)]:
4. Fill in the blank: I would use this emoji [to / for / when] ______________________

Reflected in the questions above, I operationalized how people interpret emoji along two dimensions: *sentiment* and *semantics*. Sentiment analysis involves "classifying the polarity of a given text."[5] For the purpose of my study, this meant determining whether

[3] Google provides the pure Android rendering, but many smartphone manufacturers using the Android operating system (e.g., Samsung and LG) override this rendering with their own rendering.
[4] I used the current (most updated) version of emoji for each vendor at the time of the study (December 2015).
[5] https://en.wikipedia.org/wiki/Sentiment_analysis

and how strongly the expression of a given emoji is positive, negative, or neutral. Question three elicited a numeric sentiment judgment, mirroring the -5 to 5 sentiment scale used in [59]. In the context of my study, semantics refers to what people think a given emoji means. Questions one, two, and four elicited text responses focused on semantic interpretation of the emoji.

In addition to the survey pages for the emoji in my sample, I created the same page for Apple's heart emoji (❤, U+2764). I had each participant complete this survey page twice, once at the beginning of the survey, and once at the end (after being shown their random sample of 15). This allowed me to control for quality of responses by assessing intra-rater agreement on each participant's two ratings of the heart emoji. I also assessed the variance of participants' overall ratings of the heart emoji, and found that my participants were very consistent in their sentiment evaluation: they varied, on average, by 0.54 (out of 10) sentiment points.

3.1.4 Participants

I recruited survey participants via Amazon Mechanical Turk. I required participants to be located in the United States in order to minimize interpretation differences that may arise from geographic and cultural influence, although this is an interesting direction of future work. In pilot testing my survey, I estimated that it would take roughly 30 to 35 seconds to complete each emoji survey page. Prorating from a minimum wage of $8 per hour, this equated to about $0.07 per emoji page. With 17 emoji pages per survey (random sample of 15 plus the heart emoji page shown twice), I compensated participants $1.20 for completing the survey.

My participants had a record of high quality work on Mechanical Turk: they each had at least 97% of their work approved with at least 1,000 approved tasks completed. Still, I calculated intra-rater reliability to ensure consistency within each participant's ratings. I computed the difference between each participant's pair of sentiment ratings for the heart emoji character. Out of the 334 participants, 308 (92%) of the participants differed by zero or one rating. I considered these participants to be consistent in their ratings and excluded the remaining 26 participant responses from my dataset. To identify any low-quality participant responses that were not reflected through sentiment rating inconsistency, I also read participant responses for the heart emoji questions and excluded four more participants for problematic responses (e.g., the participant used the word “devil” to describe the heart emoji). After these quality control checks, I retained the data of 304 participants for my analysis.

Of the 304 participants, 134 were male, 169 female, and 1 other. The average age was 38.6 (SD = 12; min = 19; max = 74). With regard to smartphone platform, 35% of the participants use Apple, 8% use Google/Android, 29% Samsung, 10% LG, 1% Microsoft, and the remaining 17% use others. Participants also reported their emoji usage on a scale from “Never” to “Always”: 3% said they never use emoji, 16% rarely, 45% sometimes, 27% most of the time, and 9% indicated “always”.

3.1.5 Data for Analysis

With 304 participants each completing 15 emoji interpretations, I had a total of 4,560 emoji interpretations and ended up with approximately 37 interpretations per emoji rendering (median = 37, min = 30, max = 41).

In the midst of my analysis, I discovered an error in my emoji sample. I cross-checked back with Emojipedia, the site from which I downloaded emoji images, and discovered that some of the images in my set (automatically labelled by Unicode and platform at the time of download) had been incorrectly labeled at the time of download. I accordingly examined and reorganized my survey data to ensure that I was associating participants’ interpretations with the correct emoji rendering. I ended up with incomplete data for 3 of the 25 emoji characters I sampled, so I excluded them from my analysis (U+1F614 “pensive face,” U+1F633 “flushed face,” and U+1F604 “smiling face with open mouth and smiling eyes”).

3.2 Analyses and Results

I conducted two separate analyses of the participants' interpretations: one for sentiment judgments and one for semantics, as indicated in the open-text questions. I next detail my methods and results for each analysis.

3.2.1 Sentiment Analysis

In this section, I explore the role that sentiment may play in emoji misconstrual. I describe my methods and relevant results for each of my research questions.

3.2.1.1 Methods

For each emoji rendering, I have 30 to 41 sentiment scores that are between -5 (most negative) and 5 (most positive). In order to understand the degree to which individual participants disagree on the sentiment of an emoji rendering, I computed the difference (i.e., distance) between pairs of participants' sentiment scores for that rendering. These values can range from zero (perfect agreement) to 10 (perfect disagreement) and describe the degree to which the participants disagree on the sentiment of a given rendering.

To examine the overall variation in interpretation for specific emoji renderings (RQ1), I calculated the average of these distances to generate a *within-platform sentiment misconstrual score* for each emoji rendering. This reflects the average sentiment-based misconstrual between two people. For instance, if a given symbol has a within-platform sentiment misconstrual score of 3, the sentiment ratings of this symbol would differ by 3 points (e.g., 5 and 2), on average.

To examine variation in interpretation across platforms (RQ2), I performed a similar calculation, but on pairs of participants' sentiment scores from different renderings of the same emoji character. For a given emoji character (e.g., "face with tears of joy"), and a pair of platforms (e.g., Apple and LG), I computed all pairwise distances between the two sets of sentiment ratings, and then took the average (e.g., an Apple-LG average sentiment distance). I did this for all pairs of platforms (10 pairs total), and ended up with a platform-pair average sentiment distance for each (e.g., one for Apple-LG, one for Apple-

Microsoft, one for LG-Microsoft, etc.). I then computed the grand-mean (mean of these average sentiment distances), as the *cross-platform sentiment misconstrual score*.

3.2.1.2 Results

RQ1 (Within Platform) for Sentiment

To understand the extent to which interpretation of the sentiment of each emoji rendering varies, I ranked each rendering based on the within-platform sentiment misconstrual score in descending order for each platform. I present the top three and bottom three of this ranking in Table 3.1. With an average sentiment distance of 4.40, Microsoft's rendering of "smiling face with open mouth and tightly closed eyes" has the highest disagreement. For that emoji, 44% of participants labeled it as negative and 54% labeled it as positive, indicating a clear lack of consensus. Because Microsoft's rendering has a within-platform sentiment misconstrual score of 4.40, this means participants differed by

Table 3.1: Most and Least Within-Platform Sentiment Misconstrual

	Apple	Google	Microsoft	Samsung	LG
Top 3	3.64	3.26	4.40	3.69	2.59
	3.50	2.66	2.94	2.36	2.53
	2.72	2.61	2.35	2.29	2.51
•••	•••				
Bottom 3	1.25	1.13	1.12	1.23	1.30
	0.65	1.06	1.08	1.09	1.26
	0.45	0.62	0.66	1.08	0.63
Average (SD)	1.96 (0.77)	1.79 (0.62)	1.90 (0.54)	1.84 (0.78)	1.84 (0.59)

Top three and bottom three most different in terms of sentiment.
Higher values indicate greater response variation.

4 sentiment points, on average. On the other end is the Apple rendering 😴 of "sleeping face" with an average sentiment distance of 0.45. For that emoji, 79% of participants considered it to be neutral (sentiment = 0) and all but one of the other participants judged its sentiment to be 1 or -1.

Overall, 44 of 110 renderings (40%) have a sentiment misconstrual score larger than or equal to 2, meaning that the average amount of sentiment disagreement between two people for these individual emoji renderings is 2 or more. On the other hand, only five renderings (4.5%) have a sentiment misconstrual score of 1 or less.

I also report the average sentiment misconstrual score across all emoji renderings for each platform in Table 3.1. Apple has the highest average within-platform sentiment misconstrual (1.96); Google has the lowest (1.79).

Overall, I see that even when the emoji rendering selected by the sender is exactly the same as what the recipient sees (because both sender and recipient are using the same platform), there is still plenty of sentiment misconstrual. Indeed, if I select two participants who have rated the exact same rendering, in 25% of these cases, the participants did not agree on whether the sentiment was positive, neutral, or negative. This reflects the most straightforward form of within-platform communication, and my results suggest that, even in this case, there are clear opportunities for misconstrued communication.

RQ2 (Across Platforms) for Sentiment

I now explore variance in sentiment for renderings across platforms. In Figure 3.1, I show the distribution of *platform-pair* sentiment misconstrual scores (i.e., average sentiment distances of all possible sentiment rating pairs between two platforms for a given character) for all emoji characters (each set of five renderings are shown along the x-axis in Figure 3.1). I find that approximately 41% (9 of 22) of the emoji characters have a range wider than one sentiment unit, suggesting that at least one platform's rendering of these emoji characters is different from the other platforms. For instance, the large range for "grinning face with smiling eyes" (U+1F601) reflects the very wide disagreement

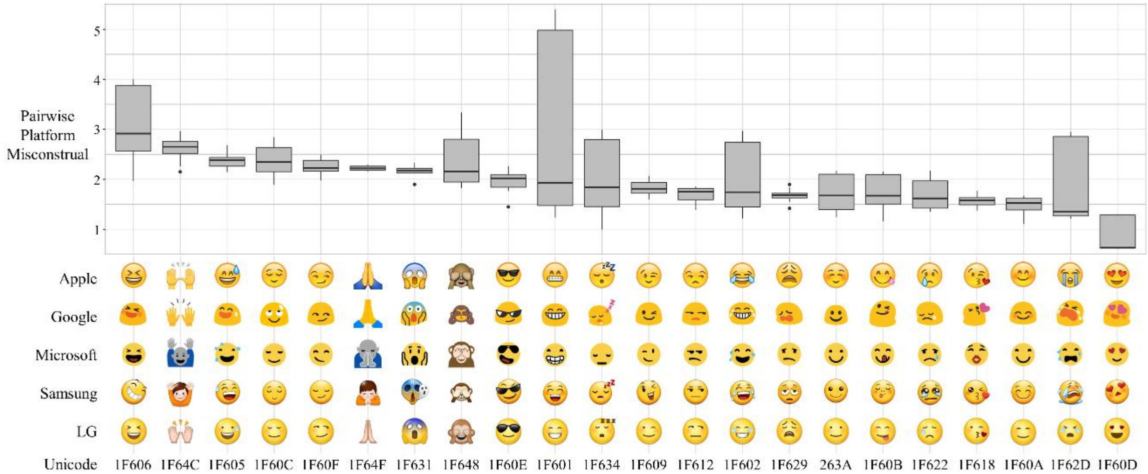


Figure 3.1: Cross-platform sentiment misconstrual scores for each emoji character. Each boxplot shows the range of platform-pair sentiment misconstrual scores. The x-axis is ordered by decreasing median platform-pair sentiment misconstrual, from left to right.

between the Apple platform and the four others (platform-pair sentiment misconstrual scores larger than 4.7), whereas the other platforms tend to agree much more among themselves (platform-pair misconstrual scores below 2). Similarly, for "sleeping face" (U+1F634), the poor agreement arises from the fact that while 91% of participants agreed that the Microsoft rendering was negative, there was a 68% chance that Samsung's rendering would be viewed as positive or neutral. It is also worth noting here that "person raising both hands in celebration" (U+1F64C) is found in the top three most different renderings for four of the five platforms, suggesting some emoji characters are simply more ambiguous than others, leading to within- and cross-platform interpretation differences.

The results from RQ1 and RQ2 regarding interpretation of sentiment suggest that there are opportunities for misconstrual of emoji in both within- and cross-platform communication contexts.

3.2.2 Semantic Analysis

Along with the perceived sentiment, differences in semantic interpretations of emoji renderings could also contribute to misconstrual.

3.2.2.1 Methods

I analyzed the free-text responses to Questions 1, 2, and 4 from my survey, which focused on the perceived meaning and use cases for the emoji. Here, I used a very similar technique to that presented above, adapted for text responses. For each participant's answer for each rendering, I aggregated their text responses to all three questions, removed stop words and stemmed word tokens (using the snowball stemmer implemented in the Scikit-Learn Python library) and then converted the text to word vectors using a standard bag-of-words model. For each rendering, I ended up with 30 to 41 word vectors representing the responses of different participants. I applied a TF-IDF transformation to all of the word vectors to reduce the importance of common words that appear in all responses, e.g., "face," "something," and "etc." To measure the difference between two participants' semantic interpretations of a given emoji rendering, I computed the cosine distance between the two participants' word vectors for that rendering. Then to measure the overall variation in interpretation for a given emoji rendering, I computed the average cosine distance of all pairs of participants' word vectors. This is similar to the within-platform sentiment misconstrual score above, so I refer to this as the *within-platform semantic misconstrual* score. These values range from zero to one, increasing as participants use a greater variety of words in their responses. Also note that these values are insensitive to the number of word vectors for each rendering.

To illustrate how the differences in word usage map to the values of average text distance, I present samples of aggregated responses in Table 3.2. The emoji rendering with smallest within-platform semantic misconstrual (0.52) was Apple's rendering 😍 of "smiling face with heart-shaped eyes." The responses for this rendering all focus heavily on the concept of "love." On the other hand, the emoji rendering with the largest within-platform semantic misconstrual (0.97) was Apple's rendering 😒 of "unamused face." The responses for this rendering show several different interpretations – "disappointment," "depressing," "unimpressed" and "suspicious."

Table 3.2: Example Participant Responses Interpreting a Given Emoji Rendering

Emoji	Avg. Text Distance	Randomly Selected Aggregated Responses for each Emoji
	(Min) 0.52	a cool kind of love cool love for when I was feeling loving but also a little chill
		I love you/this! love face I loved something someone else did or that I spotted.
		that I love something love I wanted to show I loved an idea, photo or person
		love something love something when i love something
	(Max) 0.97	Dismay, disappointed Disappointed I am dismayed or disappointed
		unimpressed unimpressed I saw, heard, or read something that I was indifferent towards
		dissapointed dissapointed dissapointment
		something depressing happened depression when something made me feel depressed

Example participant responses about the semantic meaning of a given emoji rendering and their relationship to pairwise word distance. The table includes emoji renderings with minimum and maximum average text distances in all emoji renderings.

To answer my two research questions with regard to semantic interpretation, I ran a similar analysis as the one for sentiment. I first used the within-platform semantic misconstrual score described above to answer RQ1. I also computed *cross-platform semantic misconstrual scores* of each emoji character, mirroring the computation for my sentiment analysis. For each emoji character (e.g., "face with tears of joy") and each pair of platforms (e.g., Apple and LG), I computed the pairwise word vector distances between the two sets of word vectors (one set for each platform rendering) and took the average (e.g., an Apple-LG average word vector distance for the "face with tears of joy" emoji). I then computed the grand-mean (mean of these platform-pair average word-vector distances) to get the *cross-platform semantic misconstrual score* for each emoji character.

3.2.2.2 Results

RQ1 (Within Platform) for Semantics

Shown in Table 3.3, I observe significant variation in the within-platform semantic misconstrual scores of all emoji renderings. For all five platforms, the top three renderings have a semantic misconstrual score (or average description text distance) of nearly one, indicating significantly different words used in responses from the participants for each of these renderings. Though the emoji characters with the largest misconstrual scores vary across platforms, the "smirking face" emoji (U+1F60F) appears in the top three for all platforms except Google. Only a few of the renderings (largely from Apple and Microsoft) were relatively similar, with average text distances around 0.6. These results suggest that, as with sentiment, many emoji evoke different interpretations from people.

Table 3.3: Most and Least Within-Platform Semantic Misconstrual

	Apple	Google	Microsoft	Samsung	LG
Top 3	0.97	0.97	0.96	0.96	0.96
	0.96	0.95	0.95	0.95	0.96
	0.95	0.94	0.95	0.95	0.93
•••	•••				
Bottom 3	0.73	0.75	0.64	0.72	0.73
	0.63	0.73	0.63	0.72	0.69
	0.52	0.72	0.54	0.71	0.69
Average (SD)	0.841 (0.111)	0.844 (0.078)	0.823 (0.115)	0.844 (0.080)	0.845 (0.087)

Top three and bottom three most differently described renderings.
Higher values indicate greater response variation.

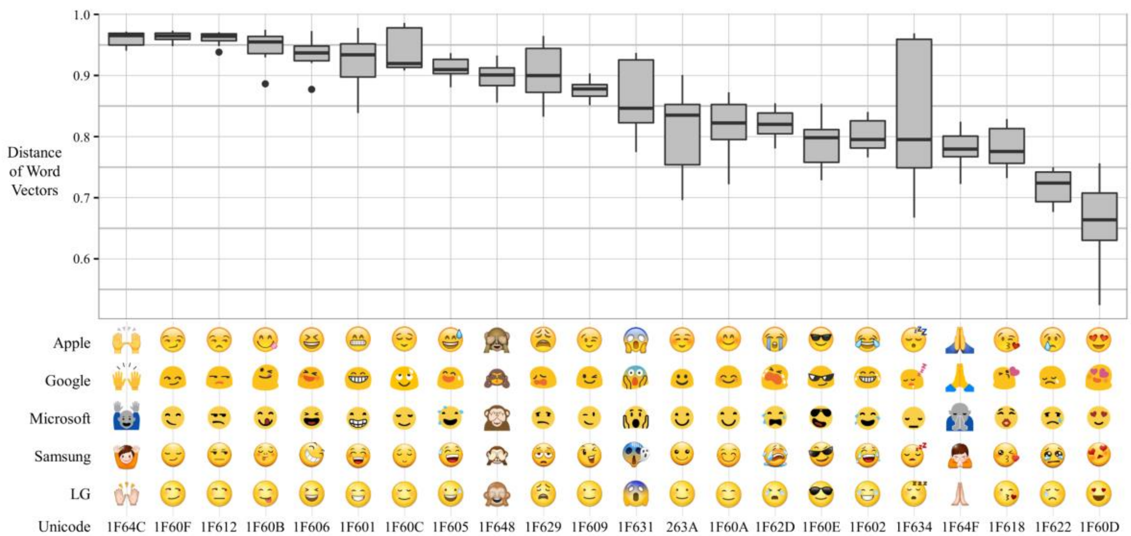


Figure 3.2: Cross-platform semantic misconstrual scores grouped by Unicode. Each boxplot shows the range of semantic misconstrual scores across the five platforms. They are ordered by decreasing median platform-pair semantic misconstrual, from left to right.

RQ2 (Across Platforms) for Semantics

Figure 3.2 shows the distribution of cross-platform semantic misconstrual scores for all platform pairs (e.g., Google and Apple, Apple and Microsoft, etc.) for all emoji characters. For each emoji character, I conducted a Kruskal-Wallis test (a non-parametric version of a one-way ANOVA, because the word vectors are not normally distributed) to explore whether the word vectors for each platform differed from one another. Indeed, I observed that there are statistically significant differences in the platform interpretations of emoji characters (Kruskal-Wallis test, p<0.001). For example, "person raising both hands in celebration" (U+1F64C) is interpreted most diversely across platforms: the top words used to describe the Apple rendering 🙌 are "hand, celebrate," "stop, clap" for the Google rendering 🙌, "praise, hand" for the LG rendering 🙌, "exciting, high" for the Microsoft rendering 🙌, and "exciting, happy" for the Samsung rendering 🙌. On the other hand, for "smiling face with heart-shaped eyes" (U+1F60D), people on all five platforms use words like "love something/someone."

It is worth pointing out that the distributions of some emoji characters have much wider variances because interpretation of a rendering for one platform largely differs from the interpretation of the renderings for the other platforms. For example, all renderings of "sleeping face" (U+1F634) except the Microsoft rendering  are clearly interpreted as a "sleeping face." In comparison, renderings of "person raising both hands in celebration" (U+1F64C) are confusing across all five platforms.

3.2.3 Results Summary

Stepping back slightly, I summarize insights from both my sentiment and my semantic findings and triangulate the degree to which both *within-platform* and *cross-platform* misconstrual may occur.

> **RQ1**: I found that in many cases, when two people consider the same emoji rendering, they may interpret both the sentiment and semantic meaning differently. In other words, there is potential for *within-platform misconstrual*. On my sentiment scale, only 4.5% of the renderings had an average misconstrual score below 1, and 40% had scores larger than 2. My semantic analysis found that very few renderings are described the same way.
>
> **RQ2**: I found that for both sentiment and semantic interpretations across platforms, there is disagreement. For a given emoji character (five renderings, one for each platform), there is clear opportunity for *cross-platform misconstrual*. 9 of the 22 (41%) emoji characters had sentiment distributions wider than one sentiment unit, and there were similar distributions of disagreement when considering how people describe renderings across platforms.

Thus, it is natural to ask: is the potential for misconstrual greater within or across platform? I found that misconstrual across platforms was incrementally larger than misconstrual within platform. More specifically, the average cross-platform sentiment and semantic misconstrual scores were 2.03 and 0.86, respectively (considering all cross-platform pairs of judgments). This is in contrast to the average within-platform sentiment and semantic misconstrual scores, which were 1.86 and 0.84, respectively (considering all within-platform pairs of judgments).

3.3 Discussion and Implications

Emoji are very popular in text communication, but I have shown that people do not interpret them in the same way. Below, I tie my results back to Clark's psycholinguistic theory of communication, presenting additional qualitative results in support of this discussion. Following that, I highlight several implications for design.

3.3.1 Contextualizing My Results in Psycholinguistic Theory

In the context of Clark's psycholinguistic theory of language use discussed above [12], let us consider the use of emoji in a hypothetical smartphone text conversation: When Abby sends an emoji, she intends a particular meaning. When Bill views the emoji, he construes or interprets what he thinks it means. If Bill's interpretation differs from Abby's intended meaning, then Bill misconstrued Abby's communication. My results suggest that people often interpret emoji in diverse fashions, potentially leading to situations like that of Abby and Bill. With discrepancy between a sender's and receiver's interpretations, the sender's intended meaning is not commonly understood by both of them, so the communication suffers. From my results, I see that this applies to emoji usage in its most simple form: within-platform communication, where the sender and the receiver see the same emoji rendering in their exchange.

Communicating across platforms, however, adds additional potential for misconstrual. Clark discusses in detail the cognition behind how people internalize communicated information. One way is through *joint personal experiences*, which fall into *joint perceptual experiences*—perception of natural signs of things—and *joint actions*—interpretation of intentional signals. Emoji usage falls into both: in addition to intending to communicate meaning, they also require perceptual interpretation to derive meaning. Clark posits that in order for a perceptual experience to be commonly understood, people must attend to—or be perceiving—the same things and become confident that they have done so in the right way. Unlike plain text where people view the same characters in their exchange, platforms effectively *translate* emoji: the emoji that the sender chose is translated to the receiver's platform's rendering. As a result, people do not attend to the same things when communicating with emoji across platform. In fact, my results show that people's interpretations for a given emoji character vary more across multiple platforms' renderings than for a single platform's rendering. This implies that communication across platforms is even more prone to misconstrual than within platform.

At the end of the survey, I asked participants if they had had any experiences with communication errors around emoji. Many participants mentioned instances in which emoji did not render on their phone (showing up as black squares), which at least informs the recipient that they are missing some meaning. However, some comments were specifically about emoji being misinterpreted in an exchange:

> *"People have interpreted the emoji meaning something different than I intended and gotten upset."* (P35)

Finally, some explicitly mention cases of miscommunication or confusion that arose from communicating across platforms:

> *"When I use an emoji on an android and my iPhone friend says that it was a sad face instead of a crying excited face."* (P179)

> *"I downloaded the new iOS platform and I sent some nice faces, and they came to my wife's phone as aliens."* (P22)

These cases provide further evidence that using emoji in communication is prone to misinterpretation, although further qualitative work would aid in understanding the broader context of this phenomenon.

3.3.2 Implications for Design

My results suggest that emoji users would benefit from convergence of emoji design across platforms. The Unicode Consortium succeeds at its goal of standardizing emoji characters such that there is a character-level mapping between platforms. However, as I have shown, this does not mean that interpretation is standardized across platforms. Converging on emoji renderings across platforms may reduce the variation of interpretation and thus lower the likelihood of miscommunication. Unfortunately, this suggestion is at odds with potential intellectual property protection of emoji renderings and the incentives of vendors to maintain distinctive branding [24].

Regardless, I also observed that a great deal of the diversity in interpretation occurs within platform, when people examine the exact same emoji rendering. One hypothesis for the mechanisms behind these results is that there is a tradeoff when it comes to "nuance" in emoji design, such as the color shade of a cheek or the slant of an eyebrow. The graphic nature of emoji affords nuanced expression, but this nuance also potentially gives rise to a greater range of interpretation. Exploring the relationship between detail and misconstrual is an important direction of future work.

Besides the design of emoji themselves, there are conceivably better ways to support emoji usage in communication. For example, when an emoji renders, smartphones could indicate whether the particular rendering being shown is the one the sender sent so the receiver can know if she is viewing the intended rendering or not. If not, smartphones could provide a way to look up the original rendering to use for interpretation rather than a translated rendering.

3.3.3 Future Work and Limitations

Though I studied 22 of the most popular anthropomorphic emoji, there are currently 2,666 total emoji characters in the Unicode Consortium standard (including non-anthropomorphic ones). Likewise, I studied 5 of the most popular mobile platform vendors, but there are at least 12 vendors with their own unique emoji renderings [75]. I also only looked at one *version* of each platform vendor's emoji even though people do not consistently use the same version of operating systems or applications. For example, emoji in Android 4.4 look different from those in Android 5.0, which look different from those in Android 6.1 (used in my study).

There are *many* different emoji renderings, and they all may be subject to differing interpretation. It would be infeasible to survey for interpretation all of them, and new ones are constantly emerging. Developing models to predict the sentiment and consistency of interpretation of a new (or unstudied) emoji is a line of research that could prove fruitful for designers and support applications that can provide feedback about the likelihood of misconstrual for a given set of renderings.

Another interesting avenue of future work lies in approaching interpretation of emoji differently. One example is the potential for culture and geography to influence differences in interpretation of emoji. Originating in Japan with global expansion, it is likely that emoji usage and interpretation is culturally dependent. Additionally, my approach to semantic analysis could be extended to use semantic relatedness measures, which would address challenges associated with vocabulary mismatch. Other ways to operationalize interpretation of emoji might be considered as well. For example, since the time of this study, Tigwell and Flatla [61] and Rodrigues et al. [52] have performed similar studies using more nuanced evaluative dimensions to capture interpretation, also finding that people varied in their interpretations.

One limitation of this work is that it considered emoji out of context (i.e., not in the presence of a larger conversation). While emoji are sometimes sent and received independently, they are most often accompanied by surrounding text (e.g., in a text message). Indeed, the following chapter describes my work addressing this limitation by exploring the variation of people's interpretations of emoji with respect to the contexts in which they appear.

3.4 Conclusion

Emoji are used alongside text in digital communication, but their visual nature leaves them open to interpretation. In addition, emoji render differently on different platforms, so people may interpret one platform's rendering differently than they interpret another platform's rendering. Psycholinguistic theory suggests that interpretation must be consistent between two people in order to avoid communication challenges. In this research, I explored whether emoji are consistently interpreted as well as whether interpretation remains consistent across renderings by different platforms. For 5 different platform renderings of 22 emoji characters, I found disagreement in terms of both sentiment and semantics, and these disagreements only increase when considering renderings across platforms.

Chapter 4

The Role of Text in Emoji Interpretation Variation

The previous chapter of my thesis work reflects interpretation of *standalone* emoji, meaning that they were interpreted in isolation. Although emoji sometimes are communicated in isolation, most often they are accompanied by surrounding text [41]. Researchers have found that emoticons can affect the interpretation of a message [37,63], but the parallel for emoji has not yet been explored, let alone the reverse relationship of text affecting the interpretation of the emoji. Other researchers have developed emoji sentiment classifiers based purely on the sentiment of text they appear in [36,45], but this reflects interpretation solely of context and not the emoji themselves.

I hypothesized that examining emoji in textual context would reduce the observed potential for miscommunication, and I conducted a study to investigate this hypothesis. Specifically, in this chapter of my thesis work, I asked:

> ***RQ****: Does the presence of text reduce inconsistencies in how emoji are interpreted, and thus the potential for miscommunication?*

I adopted an approach similar to my previous study [43] (see Section 3.1) in which I used an online survey to solicit people's interpretations of emoji. Using a between-subjects design, participants were asked to judge the sentiment expressed by emoji presented either in isolation (*standalone*) or embedded in a textual context (*in-context*).

Textual contexts were gathered by randomly selecting tweets containing the corresponding emoji character.

To investigate the hypothesis, I needed to observe and compare the potential for miscommunication associated with each condition (i.e., emoji presented standalone versus in-context). I used the same sentiment misconstrual score metric from my previous study [43] (see Section 3.2.1) to compute the potential for misconstrual of each condition. Then, I estimated the precision of these sentiment misconstrual scores via jackknifing resampling [20]. Finally, I compared the scores for each condition using Welch's t-test [64].

My results did not support my hypothesis: in general, emoji are not significantly less ambiguous when interpreted in context than when interpreted in isolation. In addition, any such differences are small relative to a baseline amount of ambiguity; roughly speaking, these differences are "just noise." Finally, my results do not trend in a particular direction: while some emoji are less ambiguous in context, others actually are *more* ambiguous in context. This work is important because it exposes that supplementing emoji usage with text does not necessarily lower the risk of miscommunication [42].

I next discuss motivation from psycholinguistic theory. Designing a robust experiment that controls for variation in types of textual contexts among other concerns was an involved process, and I outline this design following related work. I then discuss my statistical methods, followed by my results. I close by highlighting the implications of my results more broadly.

4.1 Motivation from Psycholinguistic Theory

Psycholinguist Herbert Clark's theory of language use [12] motivates studying the potential for misconstrual associated with emoji (detailed in Chapter 1), but I consulted additional theory to motivate such study of emoji in context. Bavelas and Chovil define "visible acts of meaning" as:

> "(a) [Visible acts of meaning] are sensitive to a sender-receiver relationship in that they are less likely to occur when an addressee will not see them, (b) they are analogically encoded symbols (c) their meaning can be explicated or demonstrated in context, and (d) they are fully integrated with the accompanying words." [4]

I posit that emoji are visible acts of meaning since they satisfy this definition. Bavelas and Chovil [4] argue that visible acts of meaning should be considered as a unified whole in their "Integrated Message Model." Previously these channels were often studied independently [4]. This study adopts this more "integrated" perspective by examining text and emoji together.

4.2 Survey Design

To address my research question, I conducted a survey that solicited over two thousand people's interpretations of emoji in isolation and in context. Although I borrow the basics of my experimental design from my previous study [43] (see Section 3.1), the consideration of textual context required the addition of several complex components to my survey and analytical framework. In this section, I provide an overview of my survey design, and in the next section I highlight my statistical approach. I note that both sections feature rather detailed description of methods; this is to enable my work to be replicable. I also note that while I examined both sentiment and semantic ambiguity in my previous study [43], I focused on sentiment in this study. As discussed below, considering both would have resulted in insufficient experimental power, and, as noted in the prior chapter, semantic differences have more limited interpretability.

4.2.1 Emoji and Platforms

My prior work [43] revealed variability in how people interpret emoji, identifying some as particularly subject to miscommunication. For this study, I selected the 10 emoji from

Table 4.1: Emoji, Platforms and Renderings in Study

CODE	NAME	Previous Apple	Current Apple	Previous Google	Current Google	Previous LG	Current LG	Previous Microsoft	Current Microsoft	Previous Samsung	Current Samsung	Twitter
1F606	SMILING FACE WITH OPEN MOUTH AND TIGHTLY-CLOSED EYES											
1F601	GRINNING FACE WITH SMILING EYES											
1F64C	PERSON RAISING BOTH HANDS IN CELEBRATION											
1F605	SMILING FACE WITH OPEN MOUTH AND COLD SWEAT											
1F60C	RELIEVED FACE											
1F648	SEE-NO-EVIL MONKEY											
1F64F	PERSON WITH FOLDED HANDS											
1F60F	SMIRKING FACE											
1F631	FACE SCREAMING IN FEAR											
1F602	FACE WITH TEARS OF JOY											

The 10 emoji characters (Unicode and Name) and their associated renderings for the six platforms in my study. The "Previous" column for each of the platforms shows the renderings at the time of my previous study (Winter 2015) and the "Current" column shows the current renderings at the time of this study (Fall 2016). Merged cells indicate that no changes were made to a rendering. A white background indicates inclusion in my study (all current versions and previous versions deemed to be substantively different from the updated version, 77 renderings total). A gray background indicates exclusion (previous and current versions deemed not substantively different).

that study that had the most potential for sentiment ambiguity. These "worst offenders" (see Table 4.1) are among the most frequently-used anthropomorphic emoji. Thus, by studying these ten emoji in context, I can determine whether the presence of surrounding text mitigates the problem where it is both impactful and most acute.

I considered the same five mobile platform vendors as in my previous study (Apple, Google, LG, Microsoft, and Samsung), as well as Twitter's emoji renderings (or "Twemoji") because I used Twitter as my source for text containing emoji (see the following sub-section). Importantly, all of these platforms had updated at least some of their emoji renderings since my first study. Of the five platform vendors' renderings of my 10 emoji characters (50 renderings total), 30 had been updated[6] (all 10 of Apple's renderings, 6 of Google's, 2 of LG's, all 10 of Microsoft's, and 2 of Samsung's). Some of the updates were relatively minor, for example resolution changes (particularly in

[6] According to which emoji have "changed" on each platform's page on Emojipedia. For example, http://emojipedia.org/samsung/galaxy-note-7-revised/changed/

Apple's case) and changes to adhere better to emerging emoji norms (e.g., LG's updates to match emoji skin tone norms). However, other updates involved substantial modifications in rendering appearance and effectively resulted in new implementations of the emoji characters (e.g., Microsoft's changes).

To afford comparison to my prior work while also ensuring that my results reflect the emoji state-of-the-art, I included in my study all current renderings[7] of my 10 emoji characters, as well as all previous renderings whose current renderings substantively changed relative to the prior renderings. I determined whether a rendering underwent a substantive change by having two coders (a collaborator and myself) independently assess each update as substantive or not. A substantive change was defined as having nontrivial chance of affecting one's sentiment interpretation. We achieved 87% agreement (26/30 renderings), and resolved differences jointly. In the end, 17 renderings were determined to have substantively changed. Table 4.1 shows the full set of renderings that I considered; those with white backgrounds (77 total) were included in the study.

4.2.2 Building a Corpus of Emoji Textual Contexts

I chose Twitter as a corpus for text containing emoji (i.e. emoji-bearing tweets) for two key reasons. First, Twitter is a readily available source of communication that uses emoji. Second, most tweets are public and thus more likely to be interpretable without additional hidden interpersonal context. This would not be the case, for example, in a corpus of direct sender-receiver mobile text messages as such messages are often interpreted using established norms and shared knowledge between the two parties [12,14,33], a point to which I return later. To maximize the likelihood that any participant would be able to interpret the tweets in my study (i.e., minimize the need for exogenous context), I also filtered tweets in the following ways:

- Tweets had to be written in English so that they would be readable by participants.

[7] At the time of my study, Fall of 2016

- Tweets had to be original tweets, not retweets, so they appeared in their original context.
- Tweets could not contain user mentions, to reduce the chance that they were intended for a specific individual.
- Tweets could not contain hashtags, to reduce the chance that they were intended for a particular sub-community.
- Tweets could not be from a "verified" account (i.e., celebrity or public figure), to reduce the chance that the content (and interpretation) depended on context from popular culture, current events, and other exogenous information.
- Tweets could not contain URLs or attached media (e.g., photos, video), to reduce the chance that interpretation depends on external content rather than just the surrounding text.

I used the Twitter Streaming API to randomly collect approximately 64 million public tweets between September 27 and October 15, 2016. I then filtered these tweets according to the above criteria, leaving approximately 2 million tweets to select from for my study.

To ensure that my findings about emoji in context are not tweet-specific, I randomly sampled 20 unique tweets containing each emoji character (10 x 20 = 200 tweets total) from my filtered tweet dataset. When a Twitter user crafts a tweet on a specific platform (i.e., the tweet's "source" platform), the user is interacting specifically with that platform's rendering for that emoji. Therefore, to minimize biased use cases of each emoji that may arise from differences between its renderings, I stratified the sampling of 20 tweets (for each character) to be from four identifiable rendering-specific sources. Specifically, I randomly sampled 5 tweets from each of the following[8]: (1) Twitter Web Client (originate with Twitter's emoji renderings, or Twemoji), (2) Twitter for iPhone, iPad, or Mac (originate with Apple's renderings), (3) Twitter for Android (cannot be sure

[8] For 5 of the 40 emoji-source pairs, I did not have enough tweets in my dataset due to limited data and low platform usage (Twitter Web Client and Twitter for Android), so I backfilled this deficit by pulling tweets that satisfied the same criteria from a similar dataset that was collected using the Twitter API between August and September 2015.

of the origin of emoji renderings because Android is fragmented by manufacturer, and many use their own emoji fonts), and (4) Twitter for Windows Phone (originate with Microsoft's renderings). Finally, I also made sure that each tweet contained only a single emoji.

An emoji-bearing tweet is often read on platforms that have different emoji renderings than those from platform on which the tweet was authored. For example, this tweet from my dataset was shared from an Apple device:

Will be at work in the a.m  (Apple)

But this same tweet is rendered differently for users of other platforms:

Will be at work in the a.m  (Google)

Will be at work in the a.m  (LG)

Will be at work in the a.m  (Microsoft)

Will be at work in the a.m  (Samsung)

Will be at work in the a.m  (Twitter)

This example demonstrates emoji communication across platforms, in which people see different renderings of the same emoji character in the same tweet. Even people using the same vendor may see different renderings of the same emoji if using different versions:

Will be at work in the a.m  (Current Microsoft)

Will be at work in the a.m  (Previous Microsoft)

In other words, multiple versions of a given platform's renderings essentially creates another cross-platform dimension.

To gain a cross-platform (i.e., across vendors and versions) understanding of the potential for miscommunication in using emoji with text (as I did in my previous study in using emoji without text [43]), I had to consider each sample tweet as it would be rendered on different platforms. As such, I replicated each of my 200 tweets for each rendering of the emoji contained in the tweet, as I did for the example above. In total, I gathered interpretations for 1,540 rendering-specific tweets (77 total emoji renderings x 20 tweets per rendering).

4.2.3 Experiment Design

I designed my experiment to capture the two types of data needed to make the comparison central to my research question: (1) interpretations of *standalone* emoji (replicating the work of my first study [43]) and (2) interpretations of emoji *in context*. I did this using a between-subjects experiment design; participants were randomly assigned to the standalone or context condition until the quota for each was met.

For the standalone emoji condition, I used the same survey design as my previous study [43] (see Section 3.1.3), except I collected only sentiment interpretations. I focused on sentiment interpretation because the sentiment rating scale let me precisely compare interpretations, and differences between sentiment interpretations are easier to understand than differences between open-response semantic interpretations. Importantly, considering semantics also would have affected my ability to recruit a sufficient number of participants, as the semantic component of the survey design requires a great deal more participant effort.

Participants in the *standalone* condition were randomly assigned 20 emoji renderings. Participants in the *in-context* condition were randomly assigned 20 of the emoji-containing tweets. For each tweet, I randomly showed one rendering of the emoji to display (simulating viewing the tweet on that platform-version). In both conditions, participants were instructed to judge the sentiment expressed by each emoji (standalone or in context) on an ordinal scale from Strongly Negative (-5) to Strongly Positive (5), mirroring the scale used in prior work [43,59]. For the *standalone* condition, I used the same intra-rater quality control as my previous study [43] by having each participant interpret Apple's heart emoji (❤️, U+2764) both before and after their random sample of 20 emoji. For the *in context* condition, I used "love ❤️" to show before and after the sample of tweets.

4.2.4 Participants

I recruited participants via Amazon Mechanical Turk. Since geography and culture may influence interpretation [2,46,47], I recruited only participants from the United States (limiting my findings to this cultural context); I also required participants to have 97% of their work approved with at least 1,000 approved tasks completed. I estimated it would take participants roughly 10 seconds per interpretation. With each participant providing 22 interpretations (random sample of 20 plus the heart emoji twice), I compensated each participant with $0.50 for completing the survey (prorating from a wage of $8 per hour and rounding up).

I established quotas to gather sufficient power for my statistical comparisons (see below) and to leave sufficient buffer for participants who might fail the intra-rater check. I aimed for 50 standalone evaluations of each of my 77 emoji renderings, and thus targeted 210 participants for the standalone condition and acquired 238.[9] I aimed for 30 interpretations for each of my 1,540 rendering-specific tweets, so I targeted 2,500 participants and acquired 2,356.

Following my previous study, I used intra-rater reliability results as a filter: I excluded participants whose two ratings of the Apple heart emoji differed by more than 1.0 on the sentiment scale. This eliminated 4% of the initial participant pool, leaving 235 participants in the standalone condition, and 2,247 in the context condition. Of these 2,482 participants, 1,207 identified as male, 1,269 as female, and 6 as a different gender. The median age was 33 (SD = 11; min = 18; max = 79). For emoji usage, 92 said they "Never" use emoji, 346 "Rarely," 882 "Sometimes," 737 "Frequently," and 425 "Very Frequently." 37% of participants use Apple, 31% use Samsung, 9.5% use LG, 3.6% use Google, 1.1% use Microsoft, 12.7% use other platforms, and 4.5% do not have a smartphone.

The participants from the standalone condition provided a total of 4,700 interpretations, with a median of 61interpretations per rendering (min = 58; max = 64). The participants from the context condition provided 44,903 interpretations total, with a median of 30 interpretations per rendering (mins[10] = 12,19; max = 35).

[9] This quota was exceeded because it was met after other participants had already started taking the survey.
[10] I report two minimums because the first is due to a survey flaw: one single tweet for one single rendering was not recording interpretations for about half of the survey period, until I discovered and corrected the error to start collecting data. The next least amount of interpretations per context was 19.

4.3 Analytical Methods

To measure the potential for miscommunication associated with a particular emoji in and out of textual context, I used the same metric as my previous study [43] (see Section 3.2.1): *average sentiment misconstrual score*, the average distance between all pairs of participant sentiment ratings. The motivation behind this metric is that pairwise comparisons essentially simulate communication between two people, so the greater the average distance between interpretations the more likely people are to miscommunicate. Another benefit is that this metric can be computed for a single rendering or for two different renderings of an emoji character, thus simulating both communication within and across platforms. By computing all pairwise distances between people's interpretations, I simulated the full communication space within and across platforms for the vendors and versions in my study.

I aimed to compare the variability of interpretation for when each emoji was presented standalone versus in context, and for both within- and cross-platform communication. I thus had to compute four (2x2) distinct sentiment misconstrual scores for each emoji character in my study:

- *Within-Standalone*: within-platform without textual context
- *Within-Context*: within-platform with textual context
- *Across-Standalone:* cross-platform without textual context
- *Across-Context:* cross-platform with textual context.

Within- and cross-platform computation directly follows my methods from my previous study [43]. For **within-platform** computations (with or without textual context), I computed pairwise comparisons between interpretations of the *same* emoji rendering. For **cross-platform** computations, I computed pairwise comparisons between interpretations of *different* renderings of the same emoji character (e.g., the Apple and the Google renderings). For a cross-platform misconstrual score, I first computed the score for each possible pair of platforms (e.g., Apple-Google, LG-Samsung, etc.), and then averaged across these platform-pair scores to get the overall *cross-platform* sentiment misconstrual score.

Likewise, my approach to **standalone** computations (within or across platforms) was the same as that in my previous study [43]. I computed the misconstrual score for each standalone *rendering*, and then averaged these scores to get the misconstrual score for each standalone emoji *character*. For **context** computations (within or across platforms), I computed sentiment misconstrual scores for each tweet containing a given emoji rendering, and then averaged these misconstrual scores to get the sentiment misconstrual score for each *rendering* in context. Finally, I averaged the scores for all renderings of an emoji character to get the *in-context* misconstrual score for that emoji *character*.

Misconstrual scores are not conventional statistics, so I needed to employ statistical resampling in order to estimate their precision. To do so, I used *jackknifing* resampling, which involves repeatedly re-computing the metrics with one data point removed [20]. This process allows one to estimate statistical properties (e.g., standard deviation) of arbitrarily complex metrics. Typically, a bootstrapped resample might be used in this scenario, since it is a newer and better-studied resampling method. However, in the course of my evaluation I found that bootstrapping introduces a bias when used with pairwise difference metrics like my misconstrual score. Jackknife resampling does not have this problem.

I “jackknifed” my data by participant rather than by raw sentiment scores because ratings by the same participant cannot be assumed to be independent. Also, since a participant may not have interpreted every emoji, I performed jackknife resampling individually for each emoji, where each incorporated only those participants who had interpreted the given emoji. After completing the jackknifing, I computed the standard error of the four misconstrual scores for each emoji. These standard error values allowed me to compute confidence intervals and perform statistical tests. Since my misconstrual score metric is an average (of differences), the central limit theorem implies that the metric will follow an approximately normal distribution. Therefore, I used t-distribution based confidence intervals and statistical tests.

Finally, to directly answer my research question, I compared each emoji’s standalone and context misconstrual scores, specifically *Within-Standalone* to *Within-Context*, and *Across-Standalone* to *Across-Context*. Thus, I tested the null hypothesis that the interpretation of each emoji character is equally ambiguous with or without textual context. I made these comparisons using a Welch's t-test [64], modified to use the standard error of each score (from jackknifing) instead of standard deviations divided by the square root of the sample size. Finally, because I made these comparisons for each emoji separately, I applied the Holm method [29] to adjust my *p*-values for multiple comparisons. With these adjusted *p*-values, I performed the statistical tests at a significance level of 0.05.

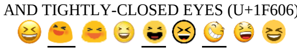
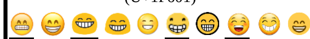
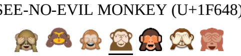
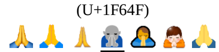
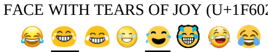
I included data for all of the 77 emoji renderings in my study (averaged across the renderings to get each emoji character's values). While this analysis combined previous and current renderings, I also repeated this analyses on the current versions of emoji characters alone, as well as on the previous versions alone (from my previous study). As I will discuss below, this analysis provided key insight into my high-level results.

4.4 Results

Table 4.2 presents the four misconstrual scores and associated 95% confidence intervals for each emoji character in the study. The "Difference" columns for the "Within" and "Across" platform conditions show the estimated *difference in misconstrual between a standalone emoji character versus the same character in textual context*. This is computed simply by subtracting each context score from the associated standalone score. If the resulting value is positive, then on average the emoji is less ambiguous in context. But if the result is negative, then on average the emoji actually is *more* ambiguous in context. Finally, I indicate the results of my hypothesis tests by highlighting in bold the differences that are statistically significant. I also display the confidence interval for each statistic.

Crucially, the lack of bold positive numbers in the "Difference" columns in Table 4.2 shows that *I found little to no support for the hypothesis that textual context reduces the potential for miscommunication when using emoji*. One emoji character – "person raising both hands in celebration" (U+1F64C) – had a significantly lower misconstrual score when considered in context (both within and across platforms, both $p<0.0001$). However,

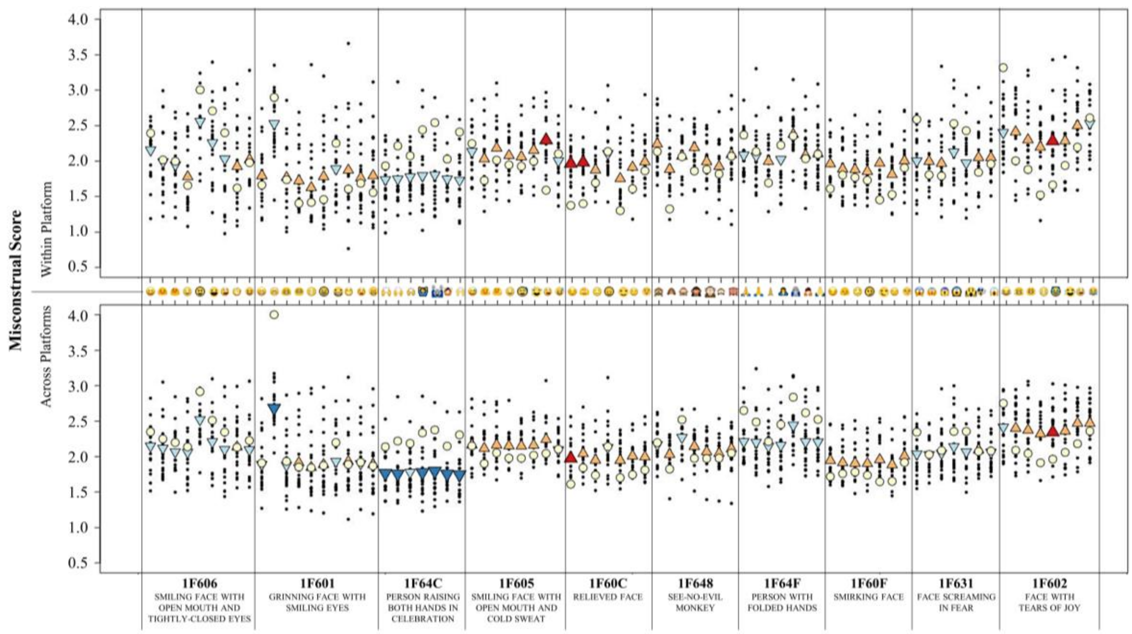
Table 4.2: Results for Comparison of Standalone and Context Sentiment Misconstrual

	WITHIN			ACROSS		
Emoji Unicode and Name	STANDALONE (Confidence Interval)	CONTEXT (Confidence Interval)	DIFFERENCE (Confidence Interval)	STANDALONE (Confidence Interval)	CONTEXT (Confidence Interval)	DIFFERENCE (Confidence Interval)
SMILING FACE WITH OPEN MOUTH AND TIGHTLY-CLOSED EYES (U+1F606) 	2.197 (2.006, 2.389)	2.074 (2.028, 2.120)	0.124 (-0.111, 0.358)	2.314 (2.145, 2.537)	2.162 (2.115, 2.209)	0.179 (-0.061, 0.419)
GRINNING FACE WITH SMILING EYES (U+1F601) 	1.769 (1.640, 1.897)	1.855 (1.813, 1.897)	-0.086 (-0.284, 0.075)	2.129 (1.994, 2.264)	1.976 (1.931, 2.020)	0.153 (-0.016, 0.323)
PERSON RAISING BOTH HANDS IN CELEBRATION (U+1F64C) 	2.235 (2.074, 2.397)	1.763 (1.718, 1.808)	**0.472*** **(0.273, 0.672)**	2.245 (2.091, 2.398)	1.767 (1.724, 1.811)	**0.477*** **(0.287, 0.668)**
SMILING FACE WITH OPEN MOUTH AND COLD SWEAT (U+1F605) 	1.944 (1.785, 2.103)	2.118 (2.071, 2.165)	-0.174 (-0.372, 0.024)	2.029 (1.874, 2.184)	2.156 (2.109, 2.202)	-0.127 (-0.319, 0.066)
RELIEVED FACE (U+1F60C) 	1.626 (1.509, 1.742)	1.941 (1.898, 1.985)	**-0.315*** **(-0.464, -0.167)**	1.799 (1.678, 1.920)	2.007 (1.963, 2.051)	-0.208 (-0.362, -0.054)
SEE-NO-EVIL MONKEY (U+1F648) 	1.879 (1.705, 2.053)	2.057 (2.010, 2.104)	-0.178 (-0.392, 0.037)	2.074 (1.894, 2.255)	2.120 (2.074, 2.166)	-0.046 (-0.268, 0.177)
PERSON WITH FOLDED HANDS (U+1F64F) 	2.129 (1.926, 2.332)	2.105 (2.056, 2.154)	0.024 (-0.225, 0.274)	2.541 (2.321, 2.761)	2.226 (2.175, 2.276)	0.315 (0.046, 0.584)
SMIRKING FACE (U+1F60F) 	1.686 (1.540, 1.833)	1.911 (1.866, 1.911)	-0.224 (-0.407, -0.042)	1.745 (1.599, 1.891)	1.932 (1.888, 1.976)	-0.187 (-0.368, -0.005)
FACE SCREAMING IN FEAR (U+1F631) 	2.135 (1.969, 2.301)	2.024 (1.970, 2.078)	0.111 (-0.097, 0.319)	2.189 (2.012, 2.367)	2.068 (2.015, 2.122)	0.121 (-0.100, 0.342)
FACE WITH TEARS OF JOY (U+1F602) 	2.142 (1.947, 2.336)	2.364 (2.314, 2.414)	-0.222 (-0.461, 0.018)	2.170 (1.961, 2.379)	2.395 (2.346, 2.444)	-0.225 (-0.481, 0.031)

The four sentiment misconstrual scores and associated confidence intervals for each emoji (renderings depicted with previous versions underlined): standalone versus in context for both within- and cross-platform analysis. The difference columns are the context scores subtracted from the standalone scores: when the value is positive, on average the emoji is less ambiguous in context, and vice versa. Differences that are bold are statistically significant at a level of 0.05; the lack of bold differences shows little support for my hypothesis.

another character – "relieved face" (U+1F60C) – has a significantly *higher* ($p < 0.001$) misconstrual score (within platform only), meaning that there is more potential for misconstrual with this emoji character when it is used with text.

Further, examining the non-significant results in Table 4.2 makes it clear that the differences between standalone and in-context misconstrual exhibit no clear directional tendency. Some emoji characters trend towards a lower misconstrual score when



Renderings occupy the x-axis.
Each • represents the misconstrual score of a tweet with the given rendering.
Each ○ represents the standalone misconstrual score of the given rendering.
A triangle represents the rendering's context misconstrual score:
▽ if less than the standalone misconstrual score and ▼ if this relationship is statistically significant (p<0.05).
△ if greater than the standalone misconstrual score and ▲ if this relationship is statistically significant (p<0.05).

Figure 4.1: Low-level visualization of misconstrual scores per emoji rendering, both within platform (top graph) and across platforms (bottom graph): The higher the point on the y-axis, the more potential there is for miscommunication, and vice versa. The variety of upward and downward pointing triangles illustrates the lack of a clear trend, in addition to the lack of statistically significant results.

considered in context; others trend towards a higher misconstrual score when considered in context.

While Table 4.2 examines the misconstrual results at the level of emoji characters, Figure 4.1 shows these results at the rendering level. The basic finding at the emoji character level also holds at the rendering level: context does not consistently reduce misconstrual. In Figure 4.1, there are 20 subgraphs: one for each of the 10 emoji characters both within and across platforms. Each subgraph depicts the misconstrual of each rendering of the given emoji character in each tweet in which it appears (each • in the figure). Each triangle represents a rendering's (average) misconstrual score in all its tweets and relates this in-context score to its standalone misconstrual score (denoted as ○): a triangle points up △ for an in-context misconstrual score greater than the standalone score (▲ for statistically significant differences), and down ▽ if it is less (▼ if significant).

If my hypothesis were supported by the data – that is, if textual context reduces emoji's potential for miscommunication – I would see a trend of ▽and ▼ triangles. But this trend is not present in Figure 4.1. Further, like the character-level results, there are few statistically significant differences.

Figure 4.1 also lets one assess visually whether any outlier tweets might be driving the results. While there are some tweets where misconstrual was much higher or lower than most tweets with a given rendering, these outliers are few.

Returning to Table 4.2, the effect sizes for the difference in misconstrual between the two conditions (i.e., the values in the "Differences" column) can be difficult to interpret in isolation, so I sought to provide context by establishing a threshold below which any differences in misconstrual can be considered negligible. To do so, I compared the values in the "Differences" column to the misconstrual score of a minimally ambiguous emoji rendering to check if any of the misconstrual differences are larger than one would expect in a minimally ambiguous context (i.e., larger than "interpretation noise"). I guessed that my control rendering ❤️ would serve as a good minimally ambiguous rendering, and this hypothesis was supported: I computed the misconstrual score for each time participants interpreted this rendering—twice when presented standalone, and twice when presented in context ("love ❤️"). This yielded four misconstrual scores for this rendering: 0.727 and 0.722 for its first and second standalone appearances, respectively, and 0.735 and 0.758 for its appearances in context. These values are all substantially below the standalone and context misconstrual values for the emoji in Table 4.2. As such, I conservatively choose 0.7 as a minimal threshold for differences in misconstrual to be considered meaningful, rather than just "interpretation noise."

Using my 0.7 threshold, I see that the effect sizes in the "Differences" column in Table 4.2 provide additional support for the conclusion that text has little to no disambiguating effect on emoji interpretation. The misconstrual differences between the standalone and context conditions, even for the few statistically significant results, are less than my “interpretation noise” threshold. Furthermore, the confidence intervals for each difference place a bound on how large of an impact context makes on emoji interpretation. None of the characters have differences that exceed the threshold of +/-0.7. In fact, more than half (12/20) of the differences are smaller than 0.4.

Finally, to understand my findings in more detail, I repeated my analysis for the previous renderings and for the current renderings. A standout result from these analyses was for the “Grinning Face with Smiling Eyes” emoji character (U+1F601). In my previous study, I found that this character had high variation in interpretation across platforms [43] and thus high potential for misconstrual, particularly due to Apple’s previous rendering  (this rendering has been substantially altered in its updated rendering; see Table 4.1). In my analysis using the previous renderings alone, I identified that there is a statistically significant reduction in the misconstrual score of this emoji character with textual context present for communication across platforms ($p < 0.01$). Rendering-level results in Figure 4.1 verify that Apple’s previous rendering is the main contributor to this effect ($p < 0.001$). This suggests that in very extreme cases, there may be support for the hypothesis that text reduces the potential for emoji-related miscommunication. I return to this point in the Discussion section below.

4.5 Discussion

My study suggests that text does not have the hypothesized disambiguation value for emoji. In this section, I discuss the implications of this finding more broadly.

An important question is *why* doesn’t text reduce emoji ambiguity? One reasonable hypothesis is that sarcasm plays a role. The survey contained an open-ended text box to gather feedback from participants, and several participants highlighted the role of sarcasm in their assessments:

> *“some of the emojis seemed sarcastic”*
>
> *“Wasn't sure how to analyze the sarcastic texts”*

Another insight as to why emoji were still ambiguous in context that was pointed out by a participant was that the texts containing the emoji were too short:

> *"A couple of the texts could use a little extra context to tell what the emoji is supposed to reflect. For instance, the "I didn't expect to see her unexpectedly" text could be either positive or negative based on context."*

With Twitter's 140 character length restriction, using tweets as the source of texts limited the amount of context accompanying emoji in the study, whereas many platforms for emoji usage are not limiting in that respect. Similarly, while using Twitter as I did (e.g., with the filtering steps outlined above) allowed me to maximize general interpretability and successfully examine general consistency of interpretation (as reflected in broadcast communication like Twitter), this approach limited the amount of *interpersonal* context (or *common ground* [12]) in the simulated communication. Future work should seek to explore emoji ambiguity in longer-form texts and in longitudinal communication in more established relationships.

Interestingly, while my study controls for the presence or absence of text to study emoji ambiguity, the reverse relationship is also worthy of examination. In other words, future work should seek to investigate whether emoji affect the ambiguity of the text they accompany. Participants reflecting in the open-text box suggested that this could be the case. For example, one participant wrote:

> *"[emoji] do have their value in that they give you a sense of security that you've gotten across the right tone in an email. Whenever I feel I need to be most clear rather than risk a misunderstanding, I insert an emoji"*

This sentiment was reflected in some qualitative responses in Cramer et al.'s [14] recent work on emoji as well.

Lastly, it is interesting to reflect on textual context's effectiveness in reducing the ambiguity of Apple's (former) rendering 😁 of the "grinning face with smiling eyes" character (U+1F601). My previous study identified a roughly bimodal distribution for sentiment interpretations for this rendering. My results from this study suggest that in these types of extreme ambiguity cases in which there are two clear senses that must be disambiguated, text may possibly help to distinguish between the two very different meanings. Examining this conjecture in detail would be a useful direction of future work.

4.5.1 Limitations

Although my study design was intentionally robust against a number of factors (e.g., idiosyncratic specific textual contexts, participant variation), it is not without limitations. First and foremost, to maximize ecological validity, I rendered the emoji images in the survey at a size that corresponds with their typical size in common use (rather than enlarged versions for easier viewing). This proved difficult for some participants that took the survey on desktop monitors. For instance, one participant wrote in an open feedback box at the end of the survey:

> *"The emojis were so small that it was difficult to determine what they were, even on a 17" monitor."*

This limitation suggests an interesting research question: how might the size of emoji affect interpretation? This could be an interesting and important direction of future work, particularly considering new ways emoji are being integrated into communication tools at different sizes. For example, in Slack and Apple Messages, when sending messages that solely contain emoji (standalone), the emoji appear larger than when you send them accompanied with text (in context).

Finally, as I mentioned above, even though I took precautions to limit the exogenous context required for interpreting tweets in my study, it is impossible to mitigate this concern entirely. For instance, some tweets may have been part of a larger series of tweets meant to be read in sequence (although the percentage of tweets in my study for which this was likely the case is very unlikely to have biased the results substantially).

4.6 Conclusion

When I found extensive variation in the interpretation of some standalone emoji in my previous study, it seemed natural that this variation would diminish, at least somewhat, if

I considered the text that often accompanies emoji. However, analyzing the results of a survey with over two thousand participants, I found little to no support for this hypothesis. In fact, the preponderance of evidence suggests that text can increase emoji ambiguity as much as it can decrease it.

Chapter 5

Effects of (Not) Seeing Emoji Rendering Differences across Platforms

My prior research from the previous chapters of my thesis found that the *multi-rendering nature of emoji* is associated with serious potential for miscommunication [43]. This potential is due to people varying from person to person in their interpretations of different renderings of the same emoji character. It is unknown, however, whether people are even aware that emoji have multiple renderings, since people can only see the emoji renderings specific to the platform they are currently using. More critically, it is also unclear if users would change their communication behavior if they could see how their messages render on other platforms. That is, we know from my prior work that people vary in their interpretations of different renderings of emoji, but we do not know if they perceive the differences between renderings to be large enough to make a difference in the context of their own communication.

Whether it is useful to give users insight into this invisible multi-rendering process and how such insight might affect their communication decisions are open questions. The stakes of these questions are significant. We know that billions of messages containing emoji are sent every day [58], and it is reasonable to expect that a non-trivial percentage of these messages are viewed on different platforms than they are sent.

The goal of this chapter of my thesis was to further our understanding of the real-world implications of the multi-rendering nature of emoji by addressing these open questions. Specifically, I first sought to explore whether people are aware of this characteristic of emoji. I next investigated whether exposing emoji rendering differences across platforms would affect communication decisions.

To accomplish this goal, I needed a way to show emoji rendering differences across platforms to people in the context of their own messages. However, no tool to do this currently exists. As such, I developed emoji rendering simulation software that parses emoji from input text and accurately displays how the text would render on a wide variety of platforms. I embedded this software in an online survey deployed on Twitter so that I could use participants' own emoji-bearing tweets to expose the multi-rendering nature of emoji. I showed participants what their tweets look like across platforms, and I surveyed whether they would have changed their tweets if they had known how their tweets appeared to followers using other platforms.

My results provide strong evidence that surfacing the multi-rendering nature of emoji would have meaningful effects on real-world text communication. At least 25% of my 710 survey respondents were not aware that emoji have multiple renderings. This suggests that a substantial proportion of people do not know that the emoji renderings they see are not always the same as the renderings their communication partners see. Additionally, 20% of my respondents indicated that they would have edited or not sent their emoji-bearing tweet if they had known how the tweet rendered on different platforms. When I generalize to the population of all tweets that include emoji, this means I estimate that millions of such potentially regretful tweets are shared per day, because people currently are not afforded visibility of emoji rendering differences across platforms.

5.1 Motivation

In human-computer interaction, invisibility of system status is considered a significant design flaw [89] and occurs when some computation (resulting in a change of state) happens that is not immediately apparent to users. Cross-platform emoji rendering creates large-scale system status invisibility issues in computer-mediated communication. However, cross-platform emoji rendering is not alone in this regard: this is also true of complex communication-mediating algorithms, like those used to curate news feeds. Indeed, in analogy to my research here but within the algorithms space, Eslami et al. [23] studied whether Facebook users were aware that their news feed was algorithmically curated. To conduct this study, they built a system known as *FeedVis* to show their participants the difference between their algorithmically curated feed and their unadulterated feed. They then used this tool to better understand users' news feed preferences.

My research stands on the shoulders of the FeedVis project. FeedVis, which is in a different application domain but targets the same visibility of system status issues, directly inspired my research questions and my primary approach: like Eslami et al. [23], I wrote software to expose a process in computer-mediated communication that was previously invisible and used that software as a probe into people's perceptions and desires with regard to the relevant communication domain.

5.2 Study Design

In this research, I operationalized my informal questions discussed above into the following formal research questions (which were motivated by those asked by Eslami et al. [23]):

1. How *aware* are people of the multi-rendering nature of emoji?
2. How do people evaluate the transformation of their communications when shown other platforms' renderings? Would people prefer to *change* their communication, given the opportunity?

My primary goal in developing my study design was to maximize both participation and ecological validity. I wanted my results to reflect a large sample of people, so I developed an online survey to collect my data.[11] Regarding ecological validity, rather

[11] An example path through the (branching) survey is included as a PDF in the appendix (Figure A.1).

than having stock examples of emoji usage and/or having to choose a small sample of emoji characters (out of 2,666) for the survey, I wanted to use participants' own communications containing emoji.

Given this goal, I decided to use Twitter as my recruitment platform because (1) emoji are very common on Twitter [45] and (2) Twitter's APIs provide automatic access to a large volume of authentic emoji use cases and their associated users. Below, I detail specifically how I recruited participants and incorporated their emoji-bearing communications (i.e., tweets) in the survey. This study was reviewed and deemed exempt from further review by my university's Institutional Review Board.

5.2.1 Twitter Recruitment

I used a recruitment approach inspired by Kariryaa et al. [32] in which potential participants whose tweets meet desired criteria (i.e., bearing emoji, in my case) are detected through the Twitter Streaming API (which returns a small random sample of all public tweets). These potential participants are then targeted (as a "tailored audience") with an advertisement that requests their participation via the Twitter ads platform.

While the core of my approach comes from Kariryaa et al. [32], I needed to adapt it for two key reasons: recency and throughput. First, I wanted the tweets seen in the survey to be as recent as possible for ecological validity purposes, so I transformed the Kariryaa et al. approach into an iterative daily cycle. Specifically, for each day during the study period, I collected potential participants for a day, then uploaded the list of participants as a tailored audience, then advertised to this tailored audience the following day, and then repeated the process. With respect to throughput, the Streaming API is an efficient means for finding tweets that match a criterion, but only when that criterion corresponds to a filter in the Streaming API. Unfortunately, there is no filter for emoji. As such, I had to instead turn to Twitter's Search API. For each emoji, I used this API to search for up to 1,000 tweets containing that emoji per day. I chose this threshold so that all 2,666 queries (corresponding to each emoji) could finish overnight, given my daily workflow. I also filtered my search so that each returned tweet met the following criteria:

- Tweet must be in English.

- Tweet cannot be a retweet (otherwise it would not be the participant's original content).
- Tweet cannot contain media. (I designed my survey to only support text tweets.)
- Tweet cannot contain user mentions. (The visualization of the tweet across all platforms would be less relevant if the participant was targeting specific people. Also, it would not be possible to infer with high accuracy the platforms of those specific people, given, e.g., vendor versions and Android variants—see below for more.)
- Tweet must be sourced from within Twitter (preventing automated tweets and spam, targeting tweets written by people).

To gain a sense of the proportions of tweets that satisfy my filter criteria, I collected all tweets returned by the Streaming API for one week during the time of the study. From this dataset, I was able to estimate that about 1.76% of tweets satisfy all of these criteria. Of these filtered tweets, approximately 38.7% bear emoji.

To recruit participants from among the Twitter users whose tweets were returned by the Search API queries, I set up an ad campaign on Twitter designed to maximize clicks to my survey link. Advertisements on Twitter are just tweets that are "promoted," so creating an advertisement is simply a matter of creating a tweet for Twitter to surface in users' feeds. Figure 5.1 shows my promoted tweet. As described above, I created new tailored audiences from the potential participants I collected daily. However, due to initially low response rates (and unpredictable and unexplained Twitter processing delays), I targeted these tailored audiences up to 10 days since they were collected. I additionally specified audience targeting criteria so that my ad was restricted to people over the age of 18 and who spoke English.

I advertised for a period of two weeks in the spring of 2018. Over this period, my advertisement received 1,316,460 impressions (views) and 6,838 link clicks (0.52%

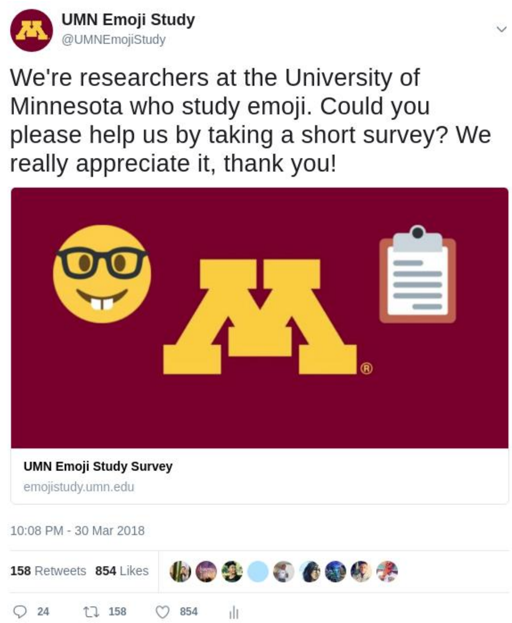


Figure 5.1: Study Advertisement Tweet

click-through rate). The cost was an average of \$0.41 per click. Of the clicks, 1,066 started the survey by providing consent and their Twitter username. 712 went on to complete the survey. The mean time from tweet to survey completion was approximately 4.5 days. I spent a total of \$2,815.73, which equates to just under \$4 per completed survey (but, as Kariryaa et al. [32] also note, this sum is paid to Twitter, not to participants).

5.2.2 Participants

I included 710 participants' surveys in my study. I removed two completed surveys: one for inappropriate open-ended responses, and one due to a small bug in the survey related to the specific device the participant was using. Of the 710 participants, 512 were female, 182 male, and 16 indicated they identified as gender non-binary (Table A.1 in the Appendix). Though the Twitter user population is disproportionately male [60], this gender distribution is somewhat expected given that women use emoji more often than men [11]. The participants were also skewed young: 75% were between 18 and 25 years old (see Table A.2 in the Appendix). This is also expected given both the populations of emoji users [90] and of Twitter users [55].

To gain a broad understanding of the geography of my respondents, I followed prior work and used the time zones attached to my respondents' Twitter accounts as a low-fidelity location indicator (e.g., [26,39,51]). From this data, it is clear that, as expected from my filters (see above), the vast majority of my respondents come from English-speaking countries. I observed that the plurality of my respondents had US/Canada time zones (e.g. "Pacific Time (US & Canada)"), and the most prominent non-US/Canada time zone was "London."

Twitter provides some indication of the "source" of each tweet in its API responses, where source is defined as the "utility used to post the tweet". Table 5.1 shows the source breakdown for tweets in the survey: 59.4% of the tweets came from Twitter for iPhone, iPad or Mac, 37.9% from Twitter for Android and 2.7% from the Twitter Web Client. Emoji renderings on Android devices are fragmented by manufacturer, but the source data given by Twitter does not capture manufacturer data. To gauge which vendors were represented in the devices respondents used to take my survey, I showed each respondent an emoji rendered natively (using the respondent's device's emoji font), and asked the

Table 5.1: Tweet Sources

Tweet Source	N	%
Twitter for iPhone / iPad / Mac	422	59.4
Twitter for Android	269	37.9%
Twitter Web Client	19	2.7%

Table 5.2: Vendors of Participants' Devices

Vendor	N	%
Apple	439	61.8%
Google (Android)	144	20.3%
Samsung	111	15.6%
Microsoft	6	0.8%
LG	4	0.6%
HTC	1	0.1%
Unknown	5	0.7%

respondent to choose the emoji rendering seen from a list of the emoji's renderings. From the answers to this question (Table 5.2), I estimate that the Twitter for Android devices are mostly split between Google's (Android) and Samsung's emoji renderings, with a very small percentage using either LG's renderings or HTC's renderings.

5.2.3 Emoji Rendering Software

Before describing the results of my survey, I first describe the software I built to simulate rendering emoji-bearing messages on different platforms. This software was used to implement the central component of my survey: asking participants whether they would edit their tweet after they could see how it appeared for followers using different platforms.

As explained above, emoji are rendered by vendor-version specific fonts. To be clear, this happens for tweets, too. Even though Twitter has its own emoji font (known as "Twemoji"), mobile Twitter applications render emoji natively using the device's emoji font. (Twemoji are used in the web client, i.e., when viewing Twitter in a browser.) Thus, since emoji-bearing tweets are often viewed on a wide variety of platforms, they are also viewed with a wide variety of renderings. When developing my emoji rendering software, I limited incorporated renderings to vendors and their associated versions that are likely to be active on Twitter (see Table A.3 in the Appendix).

The approach I took in my emoji rendering software is largely straightforward but effective: my software effectively parses out emoji characters in emoji-bearing input text and then outputs a list of HTML snippets that show how the message would render on each platform (in my sample of those are active on Twitter). Each HTML snippet includes the original text, but with the emoji character(s) replaced by emoji rendering graphic(s) (hard-coded) to show how the emoji would render on the given platform

To implement this approach, I first populated a database of emoji characters, vendors, vendor-versions, and renderings. I did this using a combination of data from the Unicode technical specification of emoji [10,91] and from Emojipedia [13,75]. To render an emoji-bearing tweet across platforms, I first used an emoji regular expression [9] to parse the emoji from the text. Then, for each vendor-version, I replaced the emoji character(s) with that vendor-version's rendering(s) and output this information as HTML.

A significant challenge in executing the above otherwise-straightforward approach centered around a particularly important and moderately-common edge case: not all vendors' versions support every emoji character, meaning that a vendor-version does not always have a rendering for a given character. This is especially an issue for older versions that do not have renderings for newer characters, but also frequently occurs when platforms implement recently-released characters at different times (the Unicode Consortium adds new characters on an annual basis). In the interest of ecological validity, when a vendor-version does not have a rendering for a given character, my software carefully adheres to the exact rules of the Unicode Technical Specification [91]. In some cases, this means rendering an "unsupported character" (□). However, in other cases, the behavior defined in the specification is more complex. In particular, some emoji characters are encoded in the Unicode standard by multiple code points, including skin-tone modified emoji, flags, families and gendered emoji (e.g., see Table 5.3). According to the specification, if an unsupported emoji character is composed of multiple code

Table 5.3: Examples of Emoji Code Points

Emoji	**Code Points and Constituent Emoji**
Beaming Face with Smiling Eyes	U+1F601
Clapping Hands: Medium-Dark Skin Tone	U+1F44F U+1F3FE [Clapping Hands] [Medium-Dark Skin Tone]
United States	U+1F1FA U+1F1F8 [Regional Indicator Symbol Letter U] [Regional Indicator Symbol Letter S]
Family: Man, Woman, Girl, Boy	U+1F468 U+200D U+1F469 U+200D U+1F467 U+200D U+1F466 [Man] [ZWJ*] [Woman] [ZWJ*] [Girl] [ZWJ*] [Boy]
Blond-Haired Woman	U+1F471 U+200D U+2640 U+FE0F [Person: Blond Hair] [ZWJ*] [Female Sign] [Variation Selector-16**]

* The Zero-Width Joiner (ZWJ) character indicates that surrounding characters should be joined into a single glyph if available.

** The Variation Selector-16 character, also known as the emoji variation selector, indicates that the preceding character be presented as an emoji (for characters that can also be presented as text, e.g., the Female Sign ♀).

points, then its component code points should be rendered individually in sequence [91] (e.g., a family might be rendered as a string of its constituent members). I implemented this approach in my rendering software. For example, refer to Figure 4.2 to see how my software rendered the emoji in Table 5.3.

5.2.4 Descriptive Statistics Regarding Emoji in Study

Each participant was shown one of her/his tweets in the survey, so I had a total of 710 tweets in my study. Of these tweets, 451 contained a single unique emoji character (either once or repeated), and 259 contained at least two different emoji characters. Across all 710 tweets, there were 1,488 total appearances of 583 unique emoji characters. Using Emojipedia's broad emoji categories [75], 164 emoji in the study were "smileys & people," 86 were "animals & nature," 34 were "food & drink," 24 were "activity," 29 were "travel & places," 47 were "objects," 63 were "symbols," 20 were "flags," and 116 were not categorized, 109 of which were skin-tone modified emoji.

As one would expect, emoji appearances in my sample followed a rough power law distribution: most emoji appeared in one (n=322) or two (n=121) tweets, with only 10 appearing in 10 or more tweets. See Table A.4 in the Appendix for the complete table of emoji in my sample.

Relatedly, the most popular emoji in my sample are also among the most popular emoji in general. Overall, though there are 2,666 total emoji characters, I estimate that the 583 emoji in my sample account for approximately 89% of all emoji appearances in tweets. This estimate is based on the distribution of emoji usage in the random sample of emoji-bearing tweets that I collected via the Twitter Streaming API as described above.

5.3 Results

In this section, I present the results from my survey, which consisted of closed-form, structured questions as well as optional, open-ended questions that inquired as to participants' reasoning behind their closed-form responses. (See Figure A.1 in the Appendix for an example run through the survey.) I primarily report descriptive quantitative statistics emerging from my structured questions. Also, though I did not employ rigorous qualitative techniques, I share insights from reading participants' open-ended responses to shed light on *possible* explanations behind my quantitative results.

5.3.1 RQ1: Awareness

I assessed participants' prior awareness of the multi-rendering nature of emoji with two questions. First, I showed the participant's tweet (rendered natively or with Twemoji, if the tweet was sourced from the Twitter web client) and asked, "Do you think that this tweet will appear exactly this way to everyone who views it?" My intention was to assess natural recall, i.e. whether participants were already aware of emoji's multi-rendering nature and had it at the top of their minds when engaging with emoji-bearing messages. I found that 47% of participants (n=334) chose "Yes," they thought the tweet would appear exactly the same way to everyone who views it. This means that 47% of participants were either unaware that emoji look different on different platforms or did not recall this fact in the context of their emoji-bearing message.

In the second question, I showed the participant's tweet and asked more explicitly, "Did you know that the emoji in your tweet will appear differently to other users on Twitter? For example, your tweet will appear as the following on the associated devices / operating systems:" and then showed the renderings of their tweet across platforms (like in Figure 4.2). For this question, 25% of participants (n=178) chose "No, I did not know this." The difference between this 25% and the 47% number from the first question can likely be interpreted as, at least in part, a manifestation of the expected effects of recognition versus recall; once prompted, some participants likely had an "oh yeah!" reaction. Some portion of this difference may also be due to observer-expectancy effects, which would not manifest in the wild.

Putting these results together, the 25% result from the second question can be considered a lower-bound estimate of the percentage of emoji-using Twitter users that are not aware of the multi-rendering nature of emoji. The 47% result from the first question can be considered an upper-bound estimate. Regardless of the precise true value, it is

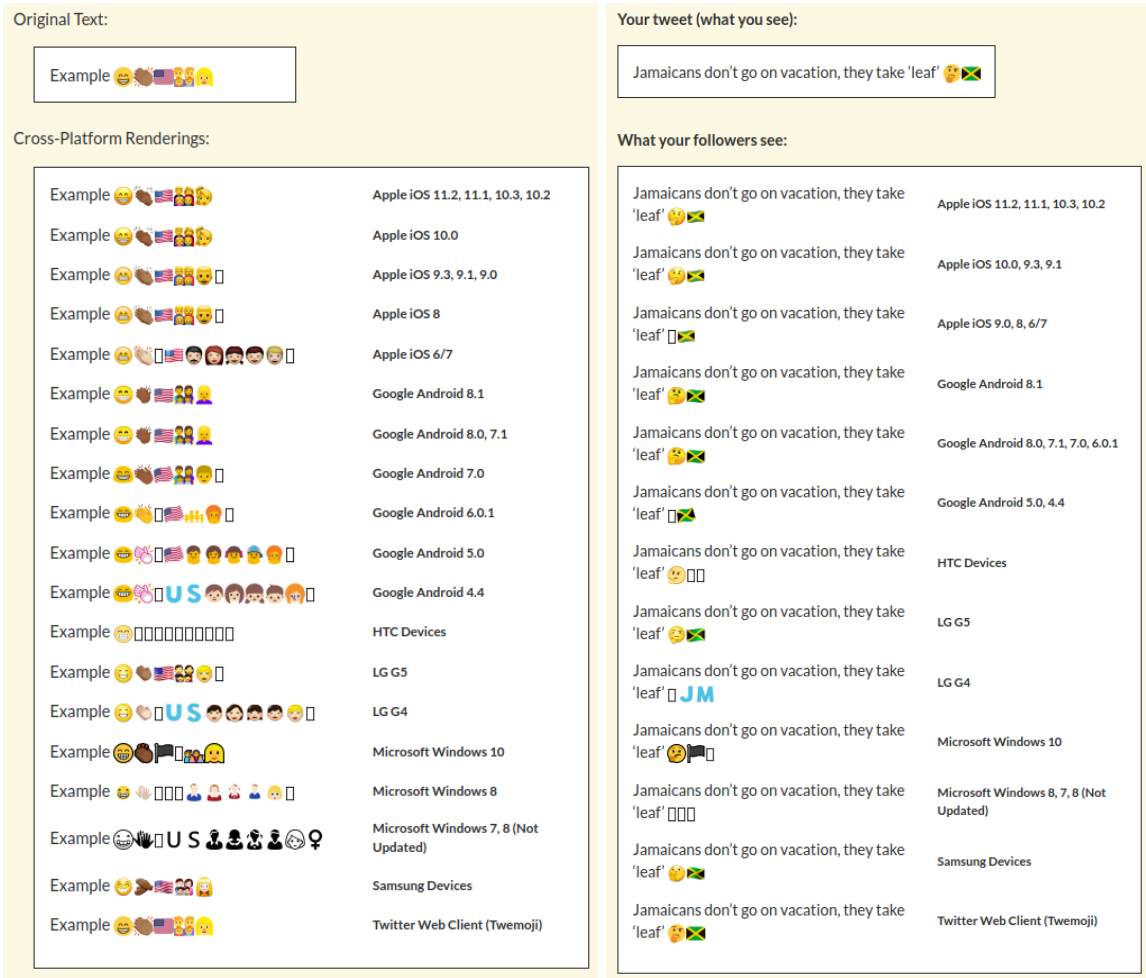


Figure 5.2: Rendering tweets across platforms. The figure on the left shows the emoji from Table 5.3 rendered across platforms by my emoji rendering software. The figure on the right shows the view in the survey of a participant's tweet rendered across platforms.

clear that *a significant proportion of Twitter users that communicate using emoji are entirely unaware that their emoji-bearing messages likely look different to many of their followers*. To put this into context, with over 300 million active Twitter users, this group of people likely contains tens of millions of people, and this does not account for those who use emoji on platforms other than Twitter.

For those in the 25% group, I provided a page in the survey that explained the multi-rendering nature of emoji and asked this group of participants about their reaction to learning this information. Looking at participant responses, some people found it moderately “interesting” or indicated that they were “surprised.” The following participant quotes reflect these sentiments:

> *“Interesting. Didn’t know how many different ways it was viewed”*

> *“I am mildly surprised there are so many types!”*

Additionally, some participants were more than just surprised and expressed shock and/or worry:

> *"I feel completely blindsided. And amazed too. I feel that it is extremely important to be aware of this because we use this platform to communicate, and if the emojis that we use are not expressed in the manner we thought it would, that might lead to misinterpretation of our statements"*

> *"Well, I was pretty bothered. What if some people misunderstood what I tweeted or posted because of the different renders of emojis? OMG 😕"*

Likewise, I also see some clearly negative responses including “*That’s annoying*,” “*Not happy*,” “*Kinda sucks*” and “*disappointed*.” Finally, the opposite end of the spectrum is also represented in participant responses. Some found it “unsurprising” or were “indifferent.” Examples of these sentiments include:

> *“Not very surprised but this is helpful”*

> *"Almost indifferent, I’m sorry for who doesn’t have an iPhone"*

> *“interesting, but it is what it is; indifferent”*

For those that were previously aware of the multi-rendering nature of emoji, I was curious about how they learned about it. Given the options of “Personal observation,” “Someone else told you,” “You read about it (e.g., in an article),” and “Other (fill in),” the vast majority (472 out of 532 aware participants) indicated that they became aware of the multi-rendering nature of emoji via personal observation. Among the other options, 16 participants (3%) became aware from someone else telling them and 27 participants (5%) became aware from reading about the multi-rendering nature of emoji. From examining participant open-ended explanations, the personal observation path consists of people who use multiple devices, who have had different devices in the past, who have seen or compared with friends' devices, and who have made inferences from seeing unsupported characters. These latter two cases are likely associated with instances of misconstrual.

In summary, with respect to RQ1, I found that at least a quarter of my participants were not aware that emoji appear (i.e., render) differently on different devices. Upon learning this information, some participants were surprised, shocked, and/or worried. For those that knew about this property of emoji, overwhelmingly it was due to personal observation.

5.3.2 RQ2: Effect on Communication Choices

After capturing data related to prior awareness, I again showed participants the renderings of their tweet across platforms, now to capture whether this would have any effect on participants' communication behavior. To first get a broad sense of the potential effect of seeing emoji rendering differences across platforms, I asked participants, "Do you think your followers' versions of the tweet convey the same message you intended to send with your tweet?" Participants could choose between the following options:

- **Yes**, I think my followers' versions convey the same message.
- I think **some** of my followers' versions convey the same message, some do not.
- **No**, I think my followers' versions do not convey the same message.

Overall, the majority (60%) of participants reported that all of the tweet renderings conveyed the same message, but a large minority 38% felt that some of them did not and 2% felt that all of the tweet renderings did not convey the same message. Some of the open responses I received from people who were among the 40% of participants for whom cross-platform emoji rendering affected the meaning of their message include:

"The emojis below are mad. Mine was meant as irritated. If I wanted it to be mad, I would've put "😤" or something sarcastic like "🙃". I hope the one reading this is using a Samsung to see my point."[12]

"Some of them are really ugly. My message is "I'm kinda pissed and mad at nothing -_- so imma just sit here stone face." The ones that don't show the defined features of the stone face (hooded eyes, eyebrows, nose, flat lips) simply does not convey MY message and possibly paints another image."

My survey next moved from interpretation to directly asking about participants' communication behavior. Specifically, following the above question, I asked: "If you had known that this is how your tweet would look to your audience, would you have sent it as-is?" Fifty-nine participants (8%) selected "No."

When asked, "How would you edit your tweet knowing this is how it looks to your audience?" participants responded as reported in Table 5.4. Table 5.4 shows that, knowing how the tweet would look across platforms, 18% of respondents (128) would have preferred to edit their tweet. These participants were relatively evenly split between choosing that they would edit the text, add more emoji, replace the emoji with another, and remove the emoji altogether.

Table 5.4: How Would Participants Edit Tweet Responses

How would you edit your tweet knowing this is how it looks to your audience?	N	%
I would not edit my tweet.	582	82.0%
I would edit the text in my tweet.	30	4.2%
I would add another emoji to my tweet.	32	4.5%
I would replace the emoji with another in my tweet.	32	4.5%
I would remove the emoji from my tweet.	29	4.1%
Other	5	0.7%

[12] I used Samsung's renderings in this quote.

Grouping the above results, 20% of tweets (144) would have been edited or not sent had the authors seen how it would look across platforms. Note that a small portion (n=16) of the participants who said they would not send their tweet as-is also selected that they would not edit their tweet. For some of these participants, multi-rendering issues likely caused their tweet to be beyond repair (other causes for this could include confusion about the question or not seeing the "Other" option).

Because I used the Twitter Search API instead of Streaming API (by necessity; see Twitter Recruitment section above), the distribution of emoji in my sample may differ somewhat from that of the population of emoji-bearing tweets. Although I observed that the most popular emoji on Twitter are also very common in my sample (see above), I wanted to formally correct for any sample-population discrepancies on this front by performing a stratified analysis of the data. Stratification is a method for adjusting the sampling weights of groups or strata within the data to account for these types of potential biases [56]. In other words, performing this analysis results in an estimate that more accurately reflects what one would expect from a true random sample of the population.

Estimating a population proportion using stratification means computing the proportion within each stratum and then multiplying it by the stratum's relative population weight. Then, the overall estimate is the sum of these weighted stratum estimates. I stratified my data by the unique combinations of emoji contained in tweets in my sample (rather than simply by each unique emoji), because some tweets contained more than one emoji character and strata cannot overlap. I estimated the relative population weight of each emoji combination by searching for tweets containing each emoji combination (using the Twitter Search API) and computing the relative popularity of that emoji combination from all of the tweets searched across all of the emoji combinations. Finally, I performed the analysis as described above, which corrected my sample proportion estimate from 20% (of emoji-bearing tweets *in my sample*) to 17.7% (of *all* emoji-bearing tweets). In other words, with this adjustment, I estimate that 17.7% of all emoji-bearing tweets would be edited or not sent as-is if the authors could have seen the emoji rendering differences across platforms.

These findings indicate that *emoji rendering differences across platforms represent a truly substantial problem for computer-mediated text communication*. To put this 17.7% figure into more context, I streamed an unfiltered sample of over 7.7 million tweets from the Streaming API to find that approximately 16.3% of all tweets contain emoji. Since there are approximately 500 million tweets shared per day [92] and approximately 16.3% contain emoji, my 17.7% estimate suggests that *there are over 14 million tweets shared per day that would not have been sent or would have been edited if the authors could have seen the emoji rendering differences across platforms.* Even if my estimate only strictly applies to my *extremely* filtered tweet population representing just 0.6% of all tweets, [13] I would still estimate that there are over 530,000 such potentially regretful tweets shared per day. However, given that emoji are also used in the rest of the population (i.e., those in other languages than English, those with media, etc.), I expect that my observed effect applies more broadly than my highly filtered context.

Hundreds of thousands if not millions of tweets *per day* is substantial, and importantly, this estimate is conservative relative to the overall real-world effect of people not being able to see emoji rendering differences across platforms. Twitter is just one of many applications that support emoji communication across platforms; others include Instagram (nearly half of all text on Instagram bear emoji [19]), Slack (8 million active daily users across 500,000 organizations [54]), and chat applications like Google Hangouts or SMS text messaging (22 billion messages are sent every day worldwide [8]). Indeed, given the increasing prevalence of emoji-bearing communication, it is not unreasonable to expect that the effect observed in this study applies to a *non-trivial percentage of all computer-mediated text communication*. However, to verify and better characterize this percentage, more research will be necessary. I articulate the research agenda that emerges from this finding in more detail in the Discussion section below.

[13] Recall that my sample contains only English, original (not retweeted), media-less, Twitter-sourced (i.e., people-written, non-spam or machine-generated) tweets containing a subset of all possible emoji. I observed that 1.76% of tweets satisfy these filter criteria, and 38.7% of these filtered tweets bear emoji (see Section 3.1). My results reflect 89% of all emoji (see Section 3.4), so, strictly speaking, altogether my sample represents 0.6% of the tweet population.

5.4 Factors Associated with Communication Choices

One question that emerges from my survey's results is why some people were concerned about how their message rendered across platforms, while others were not. I hypothesized that some factors behind this variation may include (1) characteristics of the emoji contained within the message, (2) [lack of] platform support for the emoji, (3) the role that the emoji plays in the message, and (4) the overall purpose of the message. I also hypothesized that prior awareness of the multi-rendering nature of emoji may have affected participants' communication choices. To investigate these hypotheses, I performed simple univariate statistical tests to examine whether some basic trends were present for each hypothesized factor. I also examined participant open-ended responses for evidence of whether any of these factors affected their choices.

5.4.1 Emoji-Related Factors

Emoji characters range from faces and gestures that enable people to inject nonverbal cues into text [30] to basic objects, e.g., a trophy, a basketball, a plane. Since facial expressions are nuanced and complex [27] whereas visual object recognition is simpler, I hypothesized that "face" emoji would be more liable to meaningful changes across platforms than other types of emoji. Indeed, several participants supported this hypothesis in their open responses. With respect to object recognition, one participant wrote "*It's just a trophy*" and another wrote "*the emojis are still a train and a basketball*." On the other hand, with respect to facial emoji, one participant reported "*I might be aware when I'm using other smiley emojis because some of them look really ugly in other devices…*"

The Unicode Consortium provides an emoji categorization [93] that includes several "face" categories (e.g., "face-positive," "cat-face," etc.). Using this categorization, I determined which tweets in my study contained face emoji. I observed that 24.0% of tweets that contained face emoji would be edited or not sent compared to 17.6% of tweets that did not contain any face emoji. Though not significant at the $p < 0.05$ level ($\chi 2(1,N=710) = 3.83$, $p = 0.0504$), the results suggest the expected trend: participants were more likely to indicate that they would edit or not send their tweet if it contained a face emoji.

Another emoji-related factor I hypothesized might be playing a role in my results is that, as I found in my first study [43], there are some specific emoji characters that are more likely to cause cross-platform problems (e.g., "beaming face with smiling eyes,"[14] U+1F601). Accordingly, I hypothesized that these more "ambiguous" emoji would be of more concern for participants. To investigate, I used my data for the sentiment ambiguity of the 22 emoji characters from my first study [43]. I reduced my data to tweets that only contained one of these 22 emoji characters (N=33), and I associated each tweet with its emoji's sentiment ambiguity. I observed that the ambiguity of the contained emoji was greater for tweets that would be edited or not sent (median ambiguity score[15] = 2.22) than for those that would not be edited (median ambiguity score = 1.84), but I do not find this difference to be significant at the $p < 0.05$ level (W (N=33) = 31.5, p=0.10)[16]. However, again I observed the trend that I expected to see: the more ambiguous the emoji, the more likely the participant was to prefer to edit or not send the tweet. Given my limited sample size for this test, this may be a viable hypothesis to examine in a larger study in the future.

5.4.2 Platform-Related Factors

I hypothesized that "component" (e.g., the family emoji rendered as the man, woman, girl, and boy emoji individually) or "unsupported" (e.g., □) versions of emoji characters may play a significant role in whether or not participants indicated that they would edit or not send a tweet. Participants explicitly mentioned that these phenomena were problematic, e.g.,

> *"Some people wouldn't even see the emoji. Just empty boxes."*

---

[14] Note: The Unicode has changed the name of this emoji since the time of my first study [43]

[15] This ambiguity score represents the average pairwise distance between people's sentiment interpretations of the emoji on a scale from -5 (Strongly Negative) to 5 (Strongly Positive), so the higher the ambiguity score, the higher the ambiguity [43].

[16] I use nonparametric Wilcoxon rank sum test [40] because the ambiguity measure from my first study [43] does not follow a normal distribution.

"If I'm texting someone and include an emoji that they can't see, the message may be taken a different way. I usually use emojis to lighten up a message and make it a little less serious so if they can't see if, it might change the way they read the text"

I tracked whether participants were shown "component" or "unsupported" versions of the emoji in their tweet. While these versions were more common for those that would edit or not send (80.3% vs. 78.6%), this effect was not significant ($\chi 2(1,N=710) = 0.08$, $p = 0.78$).

Somewhat unexpectedly, I saw two additional platform-related factors in participants' open-ended responses. First, several participants indicated they have a degree of "platform pride," meaning that they mentioned only really caring about the platform they use, e.g.,

"iPhone emojis are the best emojis. Everything else is just an ugly ripoff."

"Everybody knows it's the iphone emojis that are most popular and usually automatically interpret it as such."

Second, although this is very likely inaccurate, some participants felt that most of the people who would see their tweets use the same platform, e.g., "*Everyone has iphones*" and "*Because the majority of people I know have iPhones and iOS*."

5.4.3 Context-Related Factors

The specific role an emoji character plays in a tweet is also likely to be of importance with respect to whether or not someone would edit or not send the tweet. For example, is the emoji used to add new meaning to the text? To reinforce the meaning of the text? To replace text? To change the tone of the text?

I found evidence of diverse emoji roles in participant responses:

"I choose emoji to supplement, rather than convey a message"

"I usually use emojis to lighten up a message and make it a little less serious"

"The emoji was only there to make it look a little attractive."

For the purpose of this study, I captured a more general, self-reported measure of the role an emoji was playing in a tweet. Specifically, I asked participants to indicate the degree to which they agreed that their tweet needs the emoji to convey what they meant. I used a Likert scale from Strongly Disagree (-2) to Strongly Agree (2). With this information, I do not know the specific role the emoji was playing in the text, but I at least have a broad estimate of the importance of this role.

Using Wilcoxon rank sum test [40], I found that self-reported emoji importance was greater for tweets that would be edited or not sent (mean = 0.87 on Likert scale) than for those that would not be sent (mean = 0.56 on Likert scale) (W (N=710) = 35,133; p=0.013). As I expected, the more important the participant believes the emoji is to the tweet, the more likely the participant was to prefer to edit or not send the tweet.

The general purpose of the overall tweet is also likely to be important for the decision of whether or not someone would edit or not send the tweet. Mainly, how critical is it that the message be understood correctly? Who is it intended for? What would be the implications if it were misunderstood? Participants also provided comments related to these considerations. For example,

> *"Sending a tweet that's not addressed to anyone particularly is just a message you put out it's not that important to me if people on other platforms see it differently."*

> *"I just like to tweet what I want and feel at the time I don't really pay much attention as to what my followers would say or think when I tweet."*

From these assertions, emoji rendering differences across platforms appear to be less of a concern for those that use Twitter as a low-stakes communication platform. To at least partially test this hypothesis, I examined whether there was a relationship between a participant's number of followers and whether they chose to not send or edit the tweet. However, using a Wilcoxon rank sum test [40], I did not detect a significant relationship.

5.4.4 *Prior Awareness*

Finally, I hypothesized that respondents would be more likely to be affected by seeing emoji rendering differences across platforms if they were not previously aware of the multi-rendering nature of emoji. Indeed, while 15.8% of respondents who were previously aware would have edited or not sent their tweet, this number is 32.6% for people who indicated they were not previously aware of rendering differences. I found this relationship to be significant ($\chi 2(1,N=710) = 22.48$, $p < .001$).

5.5 *Conclusion*

In this chapter, I advanced the line of research that seeks to understand the impact of the multi-rendering nature of emoji. The top-level result from my survey is that emoji's potential for miscommunication identified in prior work is having demonstrable, real-world effects on people's communication. A large minority (at least 25%) of my respondents were not aware that emoji render differently across platforms, and being informed of this incited worry and frustration for some of them. I also observed that 8% of tweets in my sample *would not have been sent* had the Twitter users known how those tweets would render on viewers' platforms. Similarly, 18% of the tweets in my study would have been edited if the sender had visibility into the various potential ways the tweet would render. Indeed, my results suggest that hundreds of thousands if not millions of such potentially regretful emoji-bearing tweets are shared per day because the authors cannot see the emoji rendering differences across platforms. Moreover, this statistic represents a conservative lower-bound for the real-world effect of people not being able to see emoji rendering differences across platforms; it is likely that there are many more potentially regretful emoji-bearing messages sent or shared per day on the many applications besides Twitter that support emoji communication across platforms.

Chapter 6

Discussion

The work in this thesis suggests that emoji are ripe for miscommunication, open to interpretation in their graphic nature. Furthermore, emoji rendering differences across platforms deepen the potential for miscommunication, and my results show that seeing how emoji render across platforms would affect communication decisions in many instances of use. However, since people currently do not have this ability to see emoji rendering difference across platforms, many potentially regretful emoji-bearing messages are sent every day. A substantial proportion of people do not even know it is possible (let alone likely) that their communication partners see different emoji renderings in their exchanges.

My results suggest that emoji users would benefit from convergence of emoji design across platforms. The Unicode Consortium standardizes emoji characters such that there is a character-level mapping between platforms. This prevents characters mapping to completely different characters across platforms. However, as I have shown, this does not mean that interpretation is standardized across platforms, let alone the graphics to derive interpretation. Converging on emoji renderings across platforms may reduce the variation of interpretation and thus lower the likelihood of miscommunication. Individual vendors may be able to take steps towards convergence (e.g., by getting users to update their devices or developing single-emoji-font communication applications that are used across platforms), but unfortunately this suggestion for complete convergence across vendors is at odds with the twin forces of intellectual property law and branding incentives [24]. Thus, this is likely not a tractable solution in the foreseeable future. Additionally, I observed that a great deal of the diversity in interpretations occurs within platform, when people examine the exact same emoji rendering.

As an alternative to my recommendation to converge emoji rendering designs across platforms or to standardize to one emoji font, this section details implications for the design and development of new technologies to better support people as they communicate with emoji. It also outlines additional future work necessary to better inform this effort.

6.1 The Need for Emoji Communication Tools

A clear implication of my results is the need for new technologies to assist people with emoji communication in cross-platform environments. These technologies will likely all have the same core need as I had with my survey: to be able to simulate how an emoji-bearing message looks on other platforms. As such, in order to facilitate the development of these technologies, I am releasing the source code for my rendering software.

I can imagine many different instantiations of tools that use my rendering software to help users understand how their messages will appear to recipients. For instance, one could develop a Slack plugin that implements emoji previews, a third-party Twitter application that offers a similar functionality, or a web browser extension that surfaces the output of my rendering software for Gmail users who wish to include emoji in their emails.

These types of tools could also help to continue the trajectory of research related to emoji rendering differences across platforms that I began with this thesis work [43]. Specifically, by logging the behavior of users of these tools, one could observe how users interact with the multi-rendering nature of emoji in an even more ecologically-valid fashion than my most recent survey (from the study in Chapter 5). While my survey asked people to reflect on their own real messages, this log data would allow researchers to observe this reflection in-situ. Indeed, building one of these tools is a subject of immediate future work for my collaborators and me.

Relatedly, my results regarding factors associated with editing or not sending an emoji-bearing message suggest means by which future cross-platform emoji communication tools can be made more intelligent. For instance, one could imagine the hypothetical Slack plug-in from above popping up a warning message when a Slack message that is about to be sent is particularly vulnerable to cross-platform issues, but staying silent by default in other cases. This warning feature could apply to within-platform contexts as well, using information about the vulnerability of specific emoji renderings like that was produced by my first study.

I identified several factors that may be relevant to this prediction task, e.g., whether or not an emoji character contained paralinguistic cues, the ambiguity of the emoji character, and certain contextual properties of the emoji-bearing message. However, I do not know how these factors interact. Indeed, some factors could mediate the others. As such, implementing a feature that predicts whether a given emoji-bearing message is problematic will likely require the training of a model to understand patterns in a relatively complex decision space. To do so, much more data than was provided by my survey will be necessary. However, the log data recorded by one of the suggested tools above (or a similar tool) could likely take significant steps towards accomplishing this goal, if deployed to enough people.

Other methodological approaches could also shed more light on the factors relevant to predicting the vulnerability of a given emoji or emoji-bearing message. In particular, my results call for in-depth qualitative interview work with a limited number of participants to identify themes in what might be causing the results I observed. Specifically, what is it about a given emoji rendering that makes it more ambiguous? Then, what is it about a given set of renderings for an emoji character that together make the emoji character more ambiguous across platforms? Crucially, what is telling about how these renderings differ? Finally, what is it about the context of the message surrounding an emoji that makes it more or less concerning? The open-ended responses in my final survey highlight preliminary possibilities for these themes, but they should be verified and explored more rigorously. Further, these findings can and should also be translated into input for intelligent emoji communication tool features discussed above.

One obstacle that even such a tool would have a hard time overcoming, however, is that some of the relevant factors are challenging to capture and/or quantify at scale. This is especially true for contextual variables, e.g., the importance of the emoji to a message. More work is necessary to develop more robust and scalable measures for these factors, though some relevant work is under way. Several research efforts have contributed to understanding the different possible functions of emoji in text [14,28], and some have started trying to detect such functions automatically [44]. My results suggest that these lines of work will be useful in predicting when emoji rendering differences across platforms will have an effect.

One question is whether the family of tools suggested above would be of interest to the group of people who do not seem concerned about emoji rendering differences across platforms, e.g., survey respondents who previously knew about the multi-rendering nature of emoji but are still choosing to communicate with them (as evidenced by their tweet). The data from the survey indicates that some of these people indeed do not care about miscommunication risks because of, e.g., "platform pride." However, it is much more likely that these people do perceive miscommunication risk, but they have decided that the risk does not outweigh the benefit of the ability to communicate with emoji. Some participants alluded to this tradeoff explicitly:

> *"Your tweeps don't necessarily have the same phone as you so you try to choose the emoji that convey your thoughts or feelings as best you can with the choices you have"*

> *"I cannot predict exactly what each emoji will look like on each device so just have to keep using ones relevant to my device."*

This perspective suggests that tools that surface emoji rendering differences across platforms would also be of significant utility to those that do not on the surface seem to be concerned about these differences: such tools would enable the weighing of risk versus benefit on a per-use basis. Further, these tools would make this easy compared to the next best existing alternative of looking up each emoji character one wants to use on a website like Emojipedia. For instance, one participant wrote, "*Sometimes I wonder how it would appear on other devices, but it's too much of a hassle to check all the time so I just roll with it.*"

Similarly, my survey results suggest that such tools would also be useful in cases in which emoji rendering differences are perceived to be of limited risk. In the survey, many respondents indicated that they would send the tweet as-is after seeing the different renderings across platforms. It is likely that at least some of these participants made this choice because they decided the differences were not risky. This suggests that tools surfacing emoji renderings can provide a useful service regardless of whether the renderings are perceived to have substantial differences in the context of a message. If the differences are perceived to be risky, one can take appropriate action to edit or not send; if the differences are not perceived to be risky, one can take comfort in the decision to send as-is.

6.2 Platform Awareness Tools

My results also highlight the need for new tools to encourage what one might call *platform awareness*. Some participants assumed that emoji rendering differences across platforms were not applicable to them because they believed everyone in their audience used devices from the same vendor. This perception is likely incorrect in every case (and substantially so). However, there is currently no way for a person to assess the platform distribution in their audience short of contacting each potential recipient to ask which platform he or she is using (not to mention which version). There is also a similar risk (but in the reverse direction) in believing that all vendor-version configurations are represented in a given audience. Naturally, this is especially the case when considering direct communication (e.g., SMS, direct messages) rather than broadcasting (e.g., standard Twitter).

Unfortunately, implementing accurate platform detection is a nontrivial technical challenge. In the case of Twitter, Twitter has all the required information internally, but the company does not make this information available through its APIs (even the “source” information discussed above is far from sufficient as it does not provide information about all platforms nor versions). Because of this, my survey was not able to inform respondents about which platforms on which their tweet was actually viewed. Instead, respondents saw their tweets on the platforms that are active across all of Twitter. If platform awareness technology improves, it could be useful to replicate my work with actual per-respondent platform distribution information.

Outside of Twitter, the challenge becomes even more difficult. One way to infer platforms or devices is to analyze the User-Agent string from a browser (HTTP) request. However, there is great inconsistency in User-Agent strings, so this is very difficult without a paid service specializing in such device detection (with an expansive database of learned User-Agent strings) like Device Atlas [18]. Also, using this approach would require all audience members to make a browser request of some sort from all of their devices.

One intriguing possibility is to scale the approach I took for my survey and use emoji themselves to disambiguate platforms. A given emoji rendering reflects a vendor-version configuration. Thus, by rendering an emoji natively and then asking which from a list of renderings is the emoji being shown, the user can implicitly select the vendor and version being used to view the emoji. However, this would only be practical for certain types of applications and it would be necessary to determine and maintain a list of maximally deductive emoji renderings.

6.3 Conclusion

In conclusion, my collaborators and I identified potential for miscommunication of standalone emoji, both in terms of sentiment and semantics and both within and across platforms. I also compared the potential for sentiment miscommunication of emoji standalone versus in natural textual contexts, finding little to no support that emoji are less ambiguous in context.

After identifying the risk of miscommunicating with emoji both within and across platforms and both in isolation and in textual context, I focused on the risk due to emoji rendering differences across platforms. I estimated the real-world effect of people not being able to see emoji rendering differences across platforms by quantifying the proportion of emoji-bearing tweets whose authors would prefer to not send the tweet as-is after seeing it rendered across platforms. To do this, I developed emoji rendering simulation software that affords visibility of emoji rendering differences across platforms in the context of a given text, including when platforms do not support the given emoji. From a higher level, I also produced the first empirical information on the general awareness of the multi-rendering nature of emoji, observing that at least 25% of the Twitter users I surveyed were not aware.

Altogether this work identifies the risks of miscommunicating with emoji. The problems I have identified are important given the extraordinary popularity of emoji, but my work also informs the design and development of technology to, at least partially, mitigate these problems. The data I produced and the emoji rendering software I built can be integrated into new tools for communication applications to prevent regretful exchanges due to ambiguous emoji or emoji rendering differences across platforms. I look forward to continuing with this future work to reduce the risk of emoji-related miscommunication for the many people that use emoji.

Bibliography

1. Fathiya Al Rashdi. 2015. Forms and Functions of Emojis in Whatsapp Interaction among Omanis. Georgetown University, Washington, DC, USA.
2. Francesco Barbieri, German Kruszewski, Francesco Ronzano, and Horacio Saggion. 2016. How Cosmopolitan Are Emojis?: Exploring Emojis Usage and Meaning over Different Languages with Distributional Semantics. In *Proceedings of the 2016 ACM on Multimedia Conference* (MM '16), 531–535. https://doi.org/10.1145/2964284.2967278
3. Francesco Barbieri, Francesco Ronzano, and Horacio Saggion. 2016. What does this Emoji Mean? A Vector Space Skip-Gram Model for Twitter Emojis. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*.
4. Janet Beavin Bavelas and Nicole Chovil. 2000. Visible Acts of Meaning An Integrated Message Model of Language in Face-to-Face Dialogue. *Journal of Language and Social Psychology* 19, 2: 163–194. https://doi.org/10.1177/0261927X00019002001
5. Anshul Bhagi. Decodemoji. Retrieved August 3, 2018 from http://decodemoji.com
6. Bianca Bosker. 2014. How Emoji Get Lost In Translation. *Huffington Post*. Retrieved from https://www.huffingtonpost.com/2014/06/27/emoji-meaning_n_5530638.html
7. Jeremy Burge. 2018. Apple's Emoji Crackdown. *Emojipedia*. Retrieved July 8, 2018 from https://blog.emojipedia.org/apples-emoji-crackdown/
8. Kenneth Burke. 73 Texting Statistics That Answer All Your Questions. *Text Request*. Retrieved July 8, 2018 from https://www.textrequest.com/blog/texting-statistics-answer-questions/
9. Mathias Bynens. 2018. *emoji-regex: A regular expression to match all Emoji-only symbols as per the Unicode Standard*. Retrieved from https://github.com/mathiasbynens/emoji-regex
10. Mathias Bynens. 2018. *unicode-tr51: Emoji data extracted from Unicode Technical Report #51*. Retrieved from https://github.com/mathiasbynens/unicode-tr51
11. Zhenpeng Chen, Xuan Lu, Sheng Shen, Wei Ai, Xuanzhe Liu, and Qiaozhu Mei. 2017. Through a Gender Lens: An Empirical Study of Emoji Usage over Large-Scale Android Users. In *WWW '18 Proceedings of the 2018 World Wide Web Conference* (WWW '18), 763–772. https://doi.org/10.1145/3178876.3186157
12. H.H. Clark. 1996. *Using Language*. Cambridge University Press.
13. Ben Congdon. 2018. *python-emojipedia: :sunglasses: Emoji data from Emojipedia :tada:* Retrieved from https://github.com/bcongdon/python-emojipedia
14. Henriette Cramer, Paloma de Juan, and Joel Tetreault. 2016. Sender-intended Functions of Emojis in US Messaging. In *Proceedings of the 18th International*

Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '16), 504–509. https://doi.org/10.1145/2935334.2935370
15. Dmitry Davidov, Oren Tsur, and Ari Rappoport. 2010. Enhanced sentiment learning using twitter hashtags and smileys. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, 241–249.
16. Mark Davis and Peter Edberg. 2015. *Unicode Emoji*.
17. Daantje Derks, Agneta H. Fischer, and Arjan E.R. Bos. 2008. The role of emotion in computer-mediated communication: A review. *Computers in Human Behavior* 24, 3: 766–785. https://doi.org/10.1016/j.chb.2007.04.004
18. DeviceAtlas. Device Detection. Retrieved July 11, 2018 from https://deviceatlas.com/device-detection
19. Thomas Dimson. 2015. Emojineering Part 1: Machine Learning for Emoji Trends. *Instagram Engineering Blog*. Retrieved from http://instagram-engineering.tumblr.com/post/117889701472/emojineering-part-1-machine-learning-for-emoji
20. Bradley Efron and R.J. Tibshirani. 1994. *An Introduction to the Bootstrap*. Chapman & Hall/CRC.
21. Ben Eisner, Tim Rocktäschel, Isabelle Augenstein, Matko Bošnjak, and Sebastian Riedel. 2016. emoji2vec: Learning Emoji Representations from their Description. In *Proceedings of the 4th International Workshop on Natural Language Processing for Social Media at EMNLP 2016*.
22. Abdallah El Ali, Torben Wallbaum, Merlin Wasmann, Wilko Heuten, and Susanne CJ Boll. 2017. Face2Emoji: Using Facial Emotional Expressions to Filter Emojis. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (CHI EA '17), 1577–1584. https://doi.org/10.1145/3027063.3053086
23. Motahhare Eslami, Aimee Rickman, Kristen Vaccaro, Amirhossein Aleyasen, Andy Vuong, Karrie Karahalios, Kevin Hamilton, and Christian Sandvig. 2015. "I Always Assumed That I Wasn'T Really That Close to [Her]": Reasoning About Invisible Algorithms in News Feeds. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15), 153–162. https://doi.org/10.1145/2702123.2702556
24. Eric Goldman. 2017. *Surveying the Law of Emojis*. Social Science Research Network, Rochester, NY. Retrieved July 31, 2017 from https://papers.ssrn.com/abstract=2961060
25. Fredrik Hallsmar and Jonas Palm. 2016. Multi-class Sentiment Classification on Twitter using an Emoji Training Heuristic. KTH Royal Institute of Technology.
26. Bo Han, Afshin Rahimi, Leon Derczynski, and Timothy Baldwin. 2016. Twitter Geolocation Prediction Shared Task of the 2016 Workshop on Noisy User-generated Text. *Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT)*: 213–217.
27. Ran R. Hassin, Hillel Aviezer, and Shlomo Bentin. 2013. Inherently Ambiguous: Facial Expressions of Emotions, in Context. *Emotion Review* 5, 1: 60–65. https://doi.org/10.1177/1754073912451331

28. Susan Herring and Ashley Dainas. 2017. “Nice Picture Comment!” Graphicons in Facebook Comment Threads. *Hawaii International Conference on System Sciences 2017 (HICSS-50)*.
29. Sture Holm. 1979. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* 6, 2: 65–70.
30. Tianran Hu, Han Guo, Hao Sun, Thuy-vy Thi Nguyen, and Jiebo Luo. 2017. Spice up Your Chat: The Intentions and Sentiment Effects of Using Emoji. In *Proceedings of the 11th International AAAI Conference on Web and Social Media*.
31. David A. Huffaker and Sandra L. Calvert. 2006. Gender, Identity, and Language Use in Teenage Blogs. *Journal of Computer-Mediated Communication* 10, 2: 00–00. https://doi.org/10.1111/j.1083-6101.2005.tb00238.x
32. Ankit Kariryaa, Isaac Johnson, Johannes Schöning, and Brent Hecht. 2018. Defining and Predicting the Localness of Volunteered Geographic Information using Ground Truth Data. In *Proceedings of the 36th Annual ACM Conference on Human Factors in Computing Systems* (CHI ’18). https://doi.org/10.1145/3173574.3173839
33. Ryan Kelly and Leon Watts. 2015. Characterising the inventive appropriation of emoji as relationally meaningful in mediated close personal relationships. *Experiences of Technology Appropriation: Unanticipated Users, Usage, Circumstances, and Design*.
34. Mayu Kimura and Marie Katsurai. 2017. Automatic Construction of an Emoji Sentiment Lexicon. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017* (ASONAM ’17), 1033–1036. https://doi.org/10.1145/3110025.3110139
35. Noah Liebman and Darren Gergle. 2016. It’s (Not) Simply a Matter of Time: The Relationship Between CMC Cues and Interpersonal Affinity. In *CSCW*.
36. Kun-Lin Liu, Wu-Jun Li, and Minyi Guo. 2012. Emoticon Smoothed Language Models for Twitter Sentiment Analysis. In *AAAI’12 Proceedings of the 26th AAAI Conference on Artificial Intelligence*, 1678–1684.
37. Shao-Kang Lo. 2008. The Nonverbal Communication Functions of Emoticons in Computer-Mediated Communication. *CyberPsychology & Behavior* 11, 5: 595–597. https://doi.org/10.1089/cpb.2007.0132
38. Xuan Lu, Wei Ai, Xuanzhe Liu, Qian Li, Ning Wang, Gang Huang, and Qiaozhu Mei. 2016. Learning from the Ubiquitous Language: An Empirical Analysis of Emoji Usage of Smartphone Users. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp ’16), 770–780. https://doi.org/10.1145/2971648.2971724
39. Jalal Mahmud, Jeffrey Nichols, and Clemens Drews. 2014. Home Location Identification of Twitter Users. *ACM TIST* 5, 3: 1–21. https://doi.org/10.1145/2528548
40. H. B. Mann and D. R. Whitney. 1947. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics* 18, 1: 50–60.
41. Ben Medlock and Gretchen McCulloch. 2016. The Linguistic Secrets Found in Billions of Emoji. Retrieved October 24, 2016 from

http://www.slideshare.net/SwiftKey/the-linguistic-secrets-found-in-billions-of-emoji-sxsw-2016-presentation-59956212
42. Hannah J. Miller, Daniel Kluver, Jacob Thebault-Spieker, Loren G. Terveen, and Brent J. Hecht. 2017. Understanding Emoji Ambiguity in Context: The Role of Text in Emoji-Related Miscommunication. In *Proceedings of the 11th International AAAI Conference on Web and Social Media* (ICWSM '17), 152–161.
43. Hannah J. Miller, Jacob Thebault-Spieker, Shuo Chang, Isaac Johnson, Loren Terveen, and Brent Hecht. 2016. "Blissfully Happy" or "Ready toFight": Varying Interpretations of Emoji. In *Proceedings of the 10th International AAAI Conference on Web and Social Media* (ICWSM '16), 259–268.
44. Noa Na'aman, Hannah Provenza, and Orion Montoya. 2017. Varying Linguistic Purposes of Emoji in (Twitter) Context. *Proceedings of ACL 2017, Student Research Workshop*: 136–141.
45. Petra Kralj Novak, Jasmina Smailović, Borut Sluban, and Igor Mozetič. 2015. Sentiment of emojis. *PloS one* 10, 12.
46. Jaram Park, Young Min Baek, and Meeyoung Cha. 2014. Cross-Cultural Comparison of Nonverbal Cues in Emoticons on Twitter: Evidence from Big Data Analysis. *Journal of Communication* 64, 2: 333–354. https://doi.org/10.1111/jcom.12086
47. Jaram Park, Vladimir Barash, Clay Fink, and Meeyoung Cha. 2013. Emoticon Style: Interpreting Differences in Emoticons Across Cultures. In *International AAAI Conference on Weblogs and Social Media (ICWSM 2013)*.
48. Umashanthi Pavalanathan and Jacob Eisenstein. 2016. More emojis, less :) The competition for paralinguistic function in microblog writing. *First Monday* 21, 11. Retrieved January 11, 2017 from http://firstmonday.org/ojs/index.php/fm/article/view/6879
49. Henning Pohl, Christian Domin, and Michael Rohs. 2017. Beyond Just Text: Semantic Emoji Similarity Modeling to Support Expressive Communication 👫📲😊. *ACM Trans. Comput.-Hum. Interact.* 24, 1: 6:1–6:42. https://doi.org/10.1145/3039685
50. Henning Pohl, Dennis Stanke, and Michael Rohs. 2016. EmojiZoom: Emoji Entry via Large Overview Maps 😄🔍. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services* (MobileHCI '16), 510–517. https://doi.org/10.1145/2935334.2935382
51. Reid Priedhorsky. 2014. Inferring the Origin Locations of Tweets with Quantitative Confidence. *CSCW* 29. https://doi.org/10.1016/j.biotechadv.2011.08.021.Secreted
52. David Rodrigues, Marília Prada, Rui Gaspar, Margarida V. Garrido, and Diniz Lopes. 2018. Lisbon Emoji and Emoticon Database (LEED): Norms for emoji and emoticons in seven evaluative dimensions. *Behavior Research Methods* 50, 1: 392–405. https://doi.org/10.3758/s13428-017-0878-6
53. Samantha Scelzo. Facebook celebrates World Emoji Day by releasing some pretty impressive facts. *Mashable*. Retrieved August 2, 2018 from https://mashable.com/2017/07/17/facebook-world-emoji-day/
54. Slack. About Us. *Slack*. Retrieved July 8, 2018 from https://slack.com/about

55. Aaron Smith and Monica. 2018. Social Media Use in 2018. *Pew Research Center: Internet, Science & Tech*. Retrieved July 7, 2018 from http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/
56. T. M. F. Smith. 1991. Post-Stratification. *Journal of the Royal Statistical Society. Series D (The Statistician)* 40, 3: 315–323. https://doi.org/10.2307/2348284
57. Xavier Snelgrove. Teaching Robots to Feel: Emoji & Deep Learning. Retrieved August 2, 2018 from https://getdango.com/emoji-and-deep-learning/
58. SwiftKey. 2015. Most-used emoji revealed: Americans love skulls, Brazilians love cats, the French love hearts. *SwiftKey Blog*. Retrieved from https://blog.swiftkey.com/americans-love-skulls-brazilians-love-cats-swiftkey-emoji-meanings-report/
59. Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. Lexicon-based Methods for Sentiment Analysis. *Computational Linguistics* 37, 2: 267–307. https://doi.org/10.1162/COLI_a_00049
60. Global Twitter user distribution by gender 2018 | Statistic. *Statista*. Retrieved July 7, 2018 from https://www.statista.com/statistics/828092/distribution-of-users-on-twitter-worldwide-gender/
61. Garreth W. Tigwell and David R. Flatla. 2016. 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), 859–866. https://doi.org/10.1145/2957265.2961844
62. Chad C. Tossell, Philip Kortum, Clayton Shepard, Laura H. Barg-Walkow, Ahmad Rahmati, and Lin Zhong. 2012. A longitudinal study of emoticon use in text messaging from smartphones. *Computers in Human Behavior* 28, 2: 659–663. https://doi.org/10.1016/j.chb.2011.11.012
63. Joseph B. Walther and Kyle P. D'Addario. 2001. The impacts of emoticons on message interpretation in computer-mediated communication. *Social science computer review* 19, 3: 324–347.
64. B. L. Welch. 1947. THE GENERALIZATION OF 'STUDENT'S' PROBLEM WHEN SEVERAL DIFFERENT POPULATION VARLANCES ARE INVOLVED. *Biometrika* 34, 1–2: 28–35. https://doi.org/10.1093/biomet/34.1-2.28
65. Sanjaya Wijeratne, Lakshika Balasuriya, Amit Sheth, and Derek Doran. 2016. EmojiNet: Building a Machine Readable Sense Inventory for Emoji. In *Social Informatics*, 527–541. https://doi.org/10.1007/978-3-319-47880-7_33
66. Sanjaya Wijeratne, Lakshika Balasuriya, Amit Sheth, and Derek Doran. 2017. EmojiNet: An Open Service and API for Emoji Sense Discovery. In *Proceedings of the 11th International AAAI Conference on Web and Social Media* (ICWSM '17).
67. Peijun Zhao, Jia Jia, Yongsheng An, Jie Liang, Lexing Xie, and Jiebo Luo. 2018. Analyzing and Predicting Emoji Usages in Social Media. In *Companion Proceedings of the The Web Conference 2018* (WWW '18), 327–334. https://doi.org/10.1145/3184558.3186344
68. 2015. #ChevyGoesEmoji. *media.gm.com*. Retrieved January 24, 2018 from https://media.gm.com/content/media/us/en/chevrolet/news.detail.html/content/Pages/news/us/en/2015/jun/0622-cruze-emoji.html

69. 2015. comScore Reports July 2015 U.S. Smartphone Subscriber Market Share - comScore, Inc. Retrieved August 3, 2018 from https://www.comscore.com/Insights/Market-Rankings/comScore-Reports-July-2015-US-Smartphone-Subscriber-Market-Share
70. 2017. Chrooma Keyboard 4 adds GIF search, colored navbar, emoji suggestions, word blacklist, more. *Android Police*. Retrieved August 2, 2018 from https://www.androidpolice.com/2017/01/12/chrooma-keyboard-4-adds-gif-search-colored-navbar-emoji-suggestions-word-blacklist/
71. 2018. Slack Overhauls Emoji Support With One Catch. *Emojipedia*. Retrieved April 19, 2018 from https://blog.emojipedia.org/slack-overhauls-emoji-support/
72. 2018. Swiftmoji calling it a night, but emoji are sticking around. *SwiftKey Blog*. Retrieved August 2, 2018 from https://blog.swiftkey.com/swiftmoji-calling-night-emoji-sticking-around/
73. Oxford's 2015 Word of the Year Is This Emoji. *Time*. Retrieved January 24, 2018 from http://time.com/4114886/oxford-word-of-the-year-2015-emoji/
74. Unicode Consortium. Retrieved January 24, 2018 from http://unicode.org/consortium/consort.html
75. Emojipedia. Retrieved January 14, 2017 from http://emojipedia.org/
76. EmojiNet. Retrieved April 19, 2018 from http://emojinet.knoesis.org/home.php
77. 💬 Messenger Emoji List — All emojis for Facebook Messenger. Retrieved August 21, 2018 from https://emojipedia.org/messenger/
78. Kika Keyboard - Emoji Keyboard on the App Store. *App Store*. Retrieved August 2, 2018 from https://itunes.apple.com/us/app/kika-keyboard-emoji-keyboard/id1035199024?mt=8
79. Facemoji Emoji Keyboard:GIF, Emoji, Keyboard Theme - Apps on Google Play. Retrieved August 2, 2018 from https://play.google.com/store/apps/details?id=com.simejikeyboard&hl=en
80. Fixmoji on the App Store. *App Store*. Retrieved August 2, 2018 from https://itunes.apple.com/us/app/fixmoji/id1287040013?mt=8
81. EmoKey Joe on the App Store. *App Store*. Retrieved August 2, 2018 from https://itunes.apple.com/us/app/emokey-joe/id1144559548?mt=8
82. Gboard - the Google Keyboard - Apps on Google Play. Retrieved August 2, 2018 from https://play.google.com/store/apps/details?id=com.google.android.inputmethod.latin&hl=en_US
83. Google's Gboard can now search emoji, stickers and GIFs at once. *Engadget*. Retrieved August 2, 2018 from https://www.engadget.com/2018/02/21/gboard-beta-universal-search-new-languages/
84. Dango - Your Emoji Assistant. Retrieved August 2, 2018 from https://getdango.com
85. Use emoji on your iPhone, iPad, and iPod touch. *Apple Support*. Retrieved August 2, 2018 from https://support.apple.com/en-us/HT202332
86. EmojiTranslate. Retrieved August 3, 2018 from https://emojitranslate.com/
87. 🐌 Emoji Translator 🎲 — LingoJam. Retrieved August 3, 2018 from https://lingojam.com/%F0%9F%90%8CEmojiTranslator%F0%9F%8E%B2

88. Microsoft Windows 10 April 2018 Update Emoji List. Retrieved August 16, 2018 from https://emojipedia.org/microsoft/windows-10-april-2018-update/changed/
89. 10 Heuristics for User Interface Design: Article by Jakob Nielsen. *Nielsen Norman Group*. Retrieved April 19, 2018 from https://www.nngroup.com/articles/ten-usability-heuristics/
90. Millennials Say GIFs, Emojis Communicate Their Thoughts Better Than Words. *Time*. Retrieved April 19, 2018 from http://time.com/4834112/millennials-gifs-emojis/
91. UTS #51: Unicode Emoji. Retrieved January 24, 2018 from http://unicode.org/reports/tr51/
92. Twitter Usage Statistics - Internet Live Stats. Retrieved July 11, 2018 from http://www.internetlivestats.com/twitter-statistics/
93. Full Emoji List, v11.0. Retrieved July 10, 2018 from https://unicode.org/emoji/charts/full-emoji-list.html

Appendix

Table A.1: Participant Gender (from Chapter 5)

Gender	Twitter Impressions	Link Clicks	%	Participants	%
Female	808,861	4,347	63.6%	512	72.1%
Male	490,397	2,394	35.0%	182	25.6%
Other	17,202	97	1.4%	16	2.3%

Table A.2: Participant Age (from Chapter 5)

Age Group	N	%
18-25	534	75.21%
26-35	115	16.20%
36-45	32	4.51%
46-55	19	2.68%
56+	10	1.41%

Table A.3: Platform Versions (in Rendering Simulation Software)

Vendor	Version	Release Date	Vendor	Version	Release Date
Apple	iOS 11.2	12/2/2017	Google	Android 8.1	12/5/2017
	iOS 11.1	10/31/2017		Android 8.0	8/21/2017
	iOS 10.3	3/27/2017		Android 7.1	10/20/2016
	iOS 10.2	12/12/2016		Android 7.0	8/22/2016
	iOS 10.0	9/13/2016		Android 6.0.1	12/7/2015
	iOS 9.3	3/21/2016		Android 5.0	11/3/2014
	iOS 9.1	10/27/2015		Android 4.4	10/31/2013
	iOS 9.0	9/9/2015	Microsoft	Windows 10	10/17/2017
	iOS 8.3	4/8/2015		Windows 8.1	10/17/2013
	iOS 6.0	9/19/2012		Windows 8.0	10/26/2012
HTC	Sense 8	4/12/2016	Samsung	Experience 8.5	9/15/2017
LG	G5	4/1/2016	Twitter	Twemoji 2.5	2/22/2018
	G4	5/18/2015			

Table A.4: Emoji Used in Survey Respondents' Tweets (from Chapter 5)

Lists the emoji used in participants' tweets. The list is sorted by the total number of appearances in survey respondents' tweets. Appearance counts are given for Total occurrences (including repetitions in a given tweet), the number of unique Tweets in which the emoji appeared, and the number of unique tweets in which only that emoji appeared (Solo).

Codepoint	Emoji Name	Emojipedia Category	Appearances		
			Total	Tweet	Solo
U+1F495	Two Hearts	symbols	25	14	2
U+1F602	Face With Tears of Joy	people	22	15	1
U+1F495	Two Hearts	symbols	25	14	2
U+1F602	Face With Tears of Joy	people	22	15	1
U+2764 U+FE0F	Red Heart	symbols	20	17	0
U+1F60D	Smiling Face With Heart-Eyes	people	17	12	1
U+1F62D	Loudly Crying Face	people	15	10	1
U+1F49E	Revolving Hearts	symbols	15	10	1
U+1F496	Sparkling Heart	symbols	15	9	3
U+1F60A	Smiling Face With Smiling Eyes	people	13	11	1
U+1F497	Growing Heart	symbols	13	7	1
U+1F629	Weary Face	people	12	10	3
U+1F483 U+1F3FD	Woman Dancing: Medium Skin Tone	None	12	3	3
U+1F3B6	Musical Notes	symbols	11	11	5
U+1F631	Face Screaming in Fear	people	11	8	4
U+1F44F U+1F3FC	Clapping Hands: Medium-Light Skin Tone	None	11	6	1
U+2728	Sparkles	nature	10	10	1
U+1F622	Crying Face	people	10	6	1
U+1F648	See-No-Evil Monkey	nature	10	5	3
U+2705	White Heavy Check Mark	symbols	10	5	1
U+1F49D	Heart With Ribbon	symbols	10	4	1
U+1F6A8	Police Car Light	travel-places	10	3	2
U+1F614	Pensive Face	people	9	9	2
U+263A U+FE0F	Smiling Face	people	9	9	2
U+1F62C	Grimacing Face	people	9	7	5
U+1F634	Sleeping Face	people	9	5	1
U+1F62F	Hushed Face	people	9	4	3
U+1F44F U+1F3FD	Clapping Hands: Medium Skin Tone	None	9	3	2
U+1F47A	Goblin	people	9	2	2
U+1F914	Thinking Face	people	8	5	0
U+1F423	Hatching Chick	nature	8	4	2
U+1F61F	Worried Face	people	7	7	5
U+1F627	Anguished Face	people	7	6	4
U+1F615	Confused Face	people	7	6	4
U+1F917	Hugging Face	people	7	6	3
U+1F625	Sad but Relieved Face	people	7	6	5
U+1F644	Face With Rolling Eyes	people	7	5	1
U+1F612	Unamused Face	people	7	5	1
U+1F534	Red Circle	symbols	7	5	1
U+1F618	Face Blowing a Kiss	people	7	4	2
U+2B50	White Medium Star	nature	7	2	0
U+1F380	Ribbon	objects	7	2	0

Codepoint	Emoji Name	Emojipedia Category	Appearances		
			Total	Tweet	Solo
U+1F338	Cherry Blossom	nature	6	6	2
U+1F481 U+1F3FB U+200D U+2640 U+FE0F	Woman Tipping Hand: Light Skin Tone	None	6	6	5
U+1F61E	Disappointed Face	people	6	6	2
U+1F60B	Face Savoring Food	people	6	6	2
U+1F633	Flushed Face	people	6	6	2
U+1F609	Winking Face	people	6	6	3
U+1F937 U+200D U+2640 U+FE0F	Woman Shrugging	people	6	6	6
U+1F601	Beaming Face With Smiling Eyes	people	6	5	2
U+1F926 U+200D U+2642 U+FE0F	Man Facepalming	people	6	5	4
U+1F483	Woman Dancing	people	6	5	3
U+1F493	Beating Heart	symbols	6	5	2
U+1F5FF	Moai	objects	6	4	4
U+1F52B	Pistol	objects	6	4	1
U+1F4AA	Flexed Biceps	people	6	4	1
U+1F481	Person Tipping Hand	people	6	3	2
U+1F49F	Heart Decoration	symbols	6	3	1
U+1F44A	Oncoming Fist	people	6	2	2
U+2755	White Exclamation Mark	symbols	6	1	0
U+1F3C6	Trophy	activity	5	5	2
U+1F43B	Bear Face	nature	5	5	0
U+1F4B8	Money With Wings	objects	5	5	1
U+1F451	Crown	people	5	5	2
U+1F913	Nerd Face	people	5	5	4
U+1F62B	Tired Face	people	5	5	1
U+1F499	Blue Heart	symbols	5	5	0
U+2600 U+FE0F	Sun	nature	5	4	0
U+1F61A	Kissing Face With Closed Eyes	people	5	4	3
U+2763 U+FE0F	Heavy Heart Exclamation	symbols	5	4	2
U+1F1EF U+1F1F2	Jamaica	flags	5	3	1
U+1F44F U+1F3FE	Clapping Hands: Medium-Dark Skin Tone	None	5	3	1
U+1F635	Dizzy Face	people	5	3	2
U+1F440	Eyes	people	5	3	0
U+1F603	Grinning Face With Big Eyes	people	5	3	2
U+1F64B	Person Raising Hand	people	5	3	2
U+1F3C3 U+200D U+2642 U+FE0F	Man Running	people	5	2	1
U+1F37E	Bottle With Popping Cork	food-drink	4	4	1
U+1F343	Leaf Fluttering in Wind	nature	4	4	2
U+1F33B	Sunflower	nature	4	4	2
U+1F337	Tulip	nature	4	4	0
U+2614	Umbrella With Rain Drops	nature	4	4	0
U+1F64F U+1F3FB	Folded Hands: Light Skin Tone	None	4	4	1
U+1F918 U+1F3FB	Sign of the Horns: Light Skin Tone	None	4	4	1
U+1F926 U+1F3FB U+200D U+2640 U+FE0F	Woman Facepalming: Light Skin Tone	None	4	4	0

Codepoint	Emoji Name	Emojipedia Category	Appearances		
			Total	Tweet	Solo
U+1F630	Anxious Face With Sweat	people	4	4	3
U+1F632	Astonished Face	people	4	4	2
U+1F628	Fearful Face	people	4	4	3
U+261D U+FE0F	Index Pointing Up	people	4	4	1
U+1F922	Nauseated Face	people	4	4	0
U+1F44C	OK Hand	people	4	4	2
U+1F486	Person Getting Massage	people	4	4	2
U+1F918	Sign of the Horns	people	4	4	3
U+1F62A	Sleepy Face	people	4	4	1
U+1F641	Slightly Frowning Face	people	4	4	4
U+1F608	Smiling Face With Horns	people	4	4	1
U+1F60E	Smiling Face With Sunglasses	people	4	4	1
U+1F60F	Smirking Face	people	4	4	2
U+1F643	Upside-Down Face	people	4	4	1
U+1F49A	Green Heart	symbols	4	4	1
U+2665 U+FE0F	Heart Suit	symbols	4	4	0
U+1F3B5	Musical Note	symbols	4	4	0
U+1F370	Shortcake	food-drink	4	3	0
U+1F4A7	Droplet	nature	4	3	1
U+1F919	Call Me Hand	None	4	3	0
U+1F620	Angry Face	people	4	3	2
U+1F616	Confounded Face	people	4	3	2
U+1F62E	Face With Open Mouth	people	4	3	2
U+1F393	Graduation Cap	people	4	3	1
U+1F610	Neutral Face	people	4	3	2
U+1F445	Tongue	people	4	3	1
U+1F4AD	Thought Balloon	symbols	4	3	1
U+1F1EB U+1F1F7	France	flags	4	2	1
U+1F525	Fire	nature	4	2	0
U+1F4CD	Round Pushpin	objects	4	2	1
U+1F639	Cat Face With Tears of Joy	people	4	2	2
U+1F44F	Clapping Hands	people	4	2	1
U+1F941	Drum	activity	4	1	0
U+1F483 U+1F3FB	Woman Dancing: Light Skin Tone	None	4	1	1
U+1F3C3 U+200D U+2640 U+FE0F	Woman Running	people	4	1	1
U+26BE	Baseball	activity	3	3	0
U+1F3AC	Clapper Board	activity	3	3	1
U+1F3A7	Headphone	activity	3	3	0
U+1F3D2	Ice Hockey	activity	3	3	0
U+1F382	Birthday Cake	food-drink	3	3	1
U+1F366	Soft Ice Cream	food-drink	3	3	0
U+1F377	Wine Glass	food-drink	3	3	0
U+1F431	Cat Face	nature	3	3	1
U+1F308	Rainbow	nature	3	3	0
U+1F331	Seedling	nature	3	3	1
U+1F30A	Water Wave	nature	3	3	1
U+1F64F U+1F3FD	Folded Hands: Medium Skin Tone	None	3	3	1
U+1F44B U+1F3FB	Waving Hand: Light Skin Tone	None	3	3	3

Codepoint	Emoji Name	Emojipedia Category	Appearances		
			Total	Tweet	Solo
U+1F937 U+1F3FB U+200D U+2640 U+FE0F	Woman Shrugging: Light Skin Tone	None	3	3	2
U+1F38A	Confetti Ball	objects	3	3	1
U+1F3A5	Movie Camera	objects	3	3	2
U+1F389	Party Popper	objects	3	3	1
U+1F457	Dress	people	3	3	0
U+1F924	Drooling Face	people	3	3	0
U+1F636	Face Without Mouth	people	3	3	2
U+1F64F	Folded Hands	people	3	3	1
U+1F604	Grinning Face With Smiling Eyes	people	3	3	3
U+1F63D	Kissing Cat Face	people	3	3	2
U+1F623	Persevering Face	people	3	3	1
U+1F647	Person Bowing	people	3	3	1
U+1F621	Pouting Face	people	3	3	2
U+1F64C	Raising Hands	people	3	3	1
U+1F607	Smiling Face With Halo	people	3	3	0
U+1F927	Sneezing Face	people	3	3	2
U+1F494	Broken Heart	symbols	3	3	1
U+1F498	Heart With Arrow	symbols	3	3	1
U+1F4AF	Hundred Points	symbols	3	3	0
U+1F49C	Purple Heart	symbols	3	3	1
U+2122 U+FE0F	Trade Mark	symbols	3	3	1
U+1F49B	Yellow Heart	symbols	3	3	0
U+2708 U+FE0F	Airplane	travel-places	3	3	3
U+1F349	Watermelon	food-drink	3	2	1
U+1F42E	Cow Face	nature	3	2	1
U+1F425	Front-Facing Baby Chick	nature	3	2	0
U+1F44B U+1F3FC	Waving Hand: Medium-Light Skin Tone	None	3	2	2
U+1F489	Syringe	objects	3	2	1
U+1F91E	Crossed Fingers	people	3	2	1
U+1F63F	Crying Cat Face	people	3	2	2
U+1F637	Face With Medical Mask	people	3	2	0
U+1F923	Rolling on the Floor Laughing	people	3	2	0
U+1F44E	Thumbs Down	people	3	2	2
U+1F44B	Waving Hand	people	3	2	2
U+27A1 U+FE0F	Right Arrow	symbols	3	2	1
U+1F1EC U+1F1ED	Ghana	flags	3	1	1
U+1F1F3 U+1F1EC	Nigeria	flags	3	1	0
U+1F953	Bacon	food-drink	3	1	1
U+1F33D	Ear of Corn	food-drink	3	1	0
U+1F336 U+FE0F	Hot Pepper	food-drink	3	1	1
U+2642 U+FE0F	Male Sign	None	3	1	0
U+1F468 U+1F3FF U+200D U+1F3A4	Man Singer: Dark Skin Tone	None	3	1	1
U+1F6B6 U+1F3FF	Person Walking: Dark Skin Tone	None	3	1	0
U+1F938 U+1F3FE U+200D U+2640 U+FE0F	Woman Cartwheeling: Medium-Dark Skin Tone	None	3	1	1
U+1F485	Nail Polish	people	3	1	1

Codepoint	Emoji Name	Emojipedia Category	Appearances		
			Total	Tweet	Solo
U+2611 U+FE0F	Ballot Box With Check	symbols	3	1	1
U+1F3C0	Basketball	activity	2	2	0
U+1F3A4	Microphone	activity	2	2	2
U+1F1EE U+1F1F8	Iceland	flags	2	2	0
U+1F1EE U+1F1EA	Ireland	flags	2	2	1
U+1F1EF U+1F1F5	Japan	flags	2	2	0
U+1F3F3 U+FE0F U+200D U+1F308	Rainbow Flag	flags	2	2	1
U+1F36B	Chocolate Bar	food-drink	2	2	1
U+1F942	Clinking Glasses	food-drink	2	2	1
U+1F351	Peach	food-drink	2	2	0
U+1F34D	Pineapple	food-drink	2	2	2
U+1F35D	Spaghetti	food-drink	2	2	1
U+1F987	Bat	nature	2	2	1
U+1F327 U+FE0F	Cloud With Rain	nature	2	2	0
U+1F319	Crescent Moon	nature	2	2	2
U+1F4A8	Dashing Away	nature	2	2	0
U+1F436	Dog Face	nature	2	2	1
U+1F985	Eagle	nature	2	2	0
U+1F438	Frog Face	nature	2	2	1
U+1F30D	Globe Showing Europe-Africa	nature	2	2	2
U+1F31F	Glowing Star	nature	2	2	1
U+1F412	Monkey	nature	2	2	0
U+1F31A	New Moon Face	nature	2	2	1
U+1F419	Octopus	nature	2	2	1
U+1F334	Palm Tree	nature	2	2	1
U+1F43C	Panda Face	nature	2	2	0
U+1F427	Penguin	nature	2	2	1
U+1F437	Pig Face	nature	2	2	1
U+1F43D	Pig Nose	nature	2	2	1
U+1F407	Rabbit	nature	2	2	2
U+1F64A	Speak-No-Evil Monkey	nature	2	2	0
U+1F41A	Spiral Shell	nature	2	2	0
U+1F4A6	Sweat Droplets	nature	2	2	0
U+1F32C U+FE0F	Wind Face	nature	2	2	2
U+1F449 U+1F3FB	Backhand Index Pointing Right: Light Skin Tone	None	2	2	0
U+1F446 U+1F3FB	Backhand Index Pointing Up: Light Skin Tone	None	2	2	2
U+1F91E U+1F3FC	Crossed Fingers: Medium-Light Skin Tone	None	2	2	1
U+1F4AA U+1F3FB	Flexed Biceps: Light Skin Tone	None	2	2	1
U+1F57A U+1F3FB	Man Dancing: Light Skin Tone	None	2	2	1
U+1F926 U+1F3FC U+200D U+2642 U+FE0F	Man Facepalming: Medium-Light Skin Tone	None	2	2	0
U+1F937 U+1F3FC U+200D U+2642 U+FE0F	Man Shrugging: Medium-Light Skin Tone	None	2	2	2
U+1F44C U+1F3FB	OK Hand: Light Skin Tone	None	2	2	1
U+1F64C U+1F3FD	Raising Hands: Medium Skin Tone	None	2	2	0

Codepoint	Emoji Name	Emojipedia Category	Appearances		
			Total	Tweet	Solo
U+1F918 U+1F3FC	Sign of the Horns: Medium-Light Skin Tone	None	2	2	1
U+1F44D U+1F3FB	Thumbs Up: Light Skin Tone	None	2	2	2
U+1F44D U+1F3FD	Thumbs Up: Medium Skin Tone	None	2	2	1
U+1F44D U+1F3FE	Thumbs Up: Medium-Dark Skin Tone	None	2	2	0
U+1F926 U+1F3FC U+200D U+2640 U+FE0F	Woman Facepalming: Medium-Light Skin Tone	None	2	2	2
U+1F9DA U+1F3FB U+200D U+2640 U+FE0F	Woman Fairy: Light Skin Tone	None	2	2	0
U+1F646 U+1F3FB U+200D U+2640 U+FE0F	Woman Gesturing OK: Light Skin Tone	None	2	2	2
U+1F3C3 U+1F3FB U+200D U+2640 U+FE0F	Woman Running: Light Skin Tone	None	2	2	0
U+1F469 U+1F3FB U+200D U+1F393	Woman Student: Light Skin Tone	None	2	2	0
U+1F388	Balloon	objects	2	2	1
U+1F50B	Battery	objects	2	2	1
U+1F4F8	Camera With Flash	objects	2	2	1
U+1F48C	Love Letter	objects	2	2	1
U+1F399 U+FE0F	Studio Microphone	objects	2	2	0
U+1F4FA	Television	objects	2	2	2
U+1F47C	Baby Angel	people	2	2	1
U+1F447	Backhand Index Pointing Down	people	2	2	2
U+1F459	Bikini	people	2	2	0
U+1F613	Downcast Face With Sweat	people	2	2	0
U+1F92F	Exploding Head	people	2	2	1
U+1F611	Expressionless Face	people	2	2	1
U+1F92D	Face With Hand Over Mouth	people	2	2	1
U+1F9D0	Face With Monocle	people	2	2	2
U+1F928	Face With Raised Eyebrow	people	2	2	2
U+1F624	Face With Steam From Nose	people	2	2	0
U+1F912	Face With Thermometer	people	2	2	0
U+1F61B	Face With Tongue	people	2	2	0
U+1F463	Footprints	people	2	2	0
U+2639 U+FE0F	Frowning Face	people	2	2	1
U+1F626	Frowning Face With Open Mouth	people	2	2	1
U+1F605	Grinning Face With Sweat	people	2	2	0
U+1F606	Grinning Squinting Face	people	2	2	0
U+1F48B	Kiss Mark	people	2	2	0
U+1F937 U+200D U+2642 U+FE0F	Man Shrugging	people	2	2	2
U+1F444	Mouth	people	2	2	0
U+1F926	Person Facepalming	people	2	2	2
U+1F645	Person Gesturing No	people	2	2	1
U+1F646	Person Gesturing OK	people	2	2	1
U+1F4A9	Pile of Poo	people	2	2	1
U+1F45B	Purse	people	2	2	0

Codepoint	Emoji Name	Emojipedia Category	Appearances		
			Total	Tweet	Solo
U+1F60C	Relieved Face	people	2	2	0
U+1F48D	Ring	people	2	2	0
U+1F642	Slightly Smiling Face	people	2	2	1
U+1F929	Star-Struck	people	2	2	2
U+1F61C	Winking Face With Tongue	people	2	2	1
U+1F462	Womanâ€™s Boot	people	2	2	1
U+1F46F U+200D U+2640 U+FE0F	Women With Bunny Ears	people	2	2	2
U+1F5A4	Black Heart	symbols	2	2	1
U+274C	Cross Mark	symbols	2	2	0
U+271D U+FE0F	Latin Cross	symbols	2	2	0
U+1F9E1	Orange Heart	symbols	2	2	2
U+1F6D0	Place of Worship	symbols	2	2	0
U+2660 U+FE0F	Spade Suit	symbols	2	2	1
U+2B06 U+FE0F	Up Arrow	symbols	2	2	0
U+1F685	Bullet Train	travel-places	2	2	0
U+1F682	Locomotive	travel-places	2	2	0
U+1F69E	Mountain Railway	travel-places	2	2	0
U+26BD	Soccer Ball	activity	2	1	1
U+1F30F	Globe Showing Asia-Australia	nature	2	1	0
U+1F40D	Snake	nature	2	1	0
U+1F64C U+1F3FB	Raising Hands: Light Skin Tone	None	2	1	1
U+1F64C U+1F3FE	Raising Hands: Medium-Dark Skin Tone	None	2	1	1
U+3030 U+FE0F	Wavy Dash	None	2	1	0
U+2620 U+FE0F	Skull and Crossbones	objects	2	1	0
U+1F441 U+FE0F	Eye	people	2	1	0
U+1F64B U+200D U+2642 U+FE0F	Man Raising Hand	people	2	1	1
U+1F6B6	Person Walking	people	2	1	1
U+1F92B	Shushing Face	people	2	1	1
U+1F647 U+200D U+2640 U+FE0F	Woman Bowing	people	2	1	0
U+2B1C	White Large Square	symbols	2	1	0
U+1F949	3rd Place Medal	activity	1	1	0
U+1F47E	Alien Monster	activity	1	1	0
U+1F3C8	American Football	activity	1	1	0
U+1F3F8	Badminton	activity	1	1	1
U+1F3B3	Bowling	activity	1	1	0
U+1F94A	Boxing Glove	activity	1	1	1
U+1F3C7	Horse Racing	activity	1	1	0
U+1F939 U+200D U+2642 U+FE0F	Man Juggling	activity	1	1	0
U+1F6A3 U+200D U+2642 U+FE0F	Man Rowing Boat	activity	1	1	0
U+1F3B9	Musical Keyboard	activity	1	1	0
U+1F3BC	Musical Score	activity	1	1	0
U+1F3AD	Performing Arts	activity	1	1	0
U+1F3CA	Person Swimming	activity	1	1	1
U+1F3BB	Violin	activity	1	1	0

Codepoint	Emoji Name	Emojipedia Category	Appearances		
			Total	Tweet	Solo
U+1F93D U+200D U+2640 U+FE0F	Woman Playing Water Polo	activity	1	1	0
U+1F1E6 U+1F1FF	Azerbaijan	flags	1	1	1
U+1F1E7 U+1F1E7	Barbados	flags	1	1	0
U+1F1E8 U+1F1E6	Canada	flags	1	1	1
U+1F3C1	Chequered Flag	flags	1	1	0
U+1F1E8 U+1F1F7	Costa Rica	flags	1	1	0
U+1F1EE U+1F1F9	Italy	flags	1	1	1
U+1F1F2 U+1F1FD	Mexico	flags	1	1	0
U+1F1F3 U+1F1E6	Namibia	flags	1	1	1
U+1F1F5 U+1F1F8	Palestinian Territories	flags	1	1	1
U+1F3F4 U+E0067 U+E0062 U+E0073 U+E0063 U+E0074 U+E007F	Scotland	flags	1	1	0
U+1F1F0 U+1F1F7	South Korea	flags	1	1	1
U+1F983	Turkey	flags	1	1	1
U+1F951	Avocado	food-drink	1	1	1
U+1F32F	Burrito	food-drink	1	1	1
U+1F955	Carrot	food-drink	1	1	0
U+1F330	Chestnut	food-drink	1	1	0
U+1F37B	Clinking Beer Mugs	food-drink	1	1	0
U+1F36E	Custard	food-drink	1	1	0
U+1F346	Eggplant	food-drink	1	1	0
U+1F374	Fork and Knife	food-drink	1	1	0
U+1F354	Hamburger	food-drink	1	1	1
U+1F36F	Honey Pot	food-drink	1	1	0
U+2615	Hot Beverage	food-drink	1	1	0
U+1F95C	Peanuts	food-drink	1	1	1
U+1F350	Pear	food-drink	1	1	0
U+1F355	Pizza	food-drink	1	1	0
U+1F37F	Popcorn	food-drink	1	1	0
U+1F372	Pot of Food	food-drink	1	1	1
U+1F944	Spoon	food-drink	1	1	0
U+1F375	Teacup Without Handle	food-drink	1	1	1
U+1F379	Tropical Drink	food-drink	1	1	1
U+1F943	Tumbler Glass	food-drink	1	1	1
U+1F426	Bird	nature	1	1	0
U+1F33C	Blossom	nature	1	1	1
U+1F490	Bouquet	nature	1	1	1
U+1F408	Cat	nature	1	1	0
U+1F414	Chicken	nature	1	1	0
U+1F4A5	Collision	nature	1	1	0
U+1F40A	Crocodile	nature	1	1	0
U+1F4AB	Dizzy	nature	1	1	0
U+1F415	Dog	nature	1	1	1
U+1F54A U+FE0F	Dove	nature	1	1	0
U+1F409	Dragon	nature	1	1	0
U+1F432	Dragon Face	nature	1	1	0
U+1F418	Elephant	nature	1	1	0
U+1F411	Ewe	nature	1	1	0

Codepoint	Emoji Name	Emojipedia Category	Appearances		
			Total	Tweet	Solo
U+1F313	First Quarter Moon	nature	1	1	1
U+1F31B	First Quarter Moon Face	nature	1	1	1
U+1F41F	Fish	nature	1	1	0
U+1F340	Four Leaf Clover	nature	1	1	0
U+1F315	Full Moon	nature	1	1	0
U+1F410	Goat	nature	1	1	0
U+1F33A	Hibiscus	nature	1	1	0
U+1F41D	Honeybee	nature	1	1	0
U+1F41E	Lady Beetle	nature	1	1	0
U+1F981	Lion Face	nature	1	1	0
U+1F43E	Paw Prints	nature	1	1	0
U+1F430	Rabbit Face	nature	1	1	0
U+1F413	Rooster	nature	1	1	0
U+1F339	Rose	nature	1	1	1
U+2618 U+FE0F	Shamrock	nature	1	1	0
U+1F988	Shark	nature	1	1	0
U+1F33E	Sheaf of Rice	nature	1	1	1
U+1F40C	Snail	nature	1	1	1
U+2744 U+FE0F	Snowflake	nature	1	1	0
U+1F433	Spouting Whale	nature	1	1	0
U+1F991	Squid	nature	1	1	0
U+26C5	Sun Behind Cloud	nature	1	1	1
U+1F38B	Tanabata Tree	nature	1	1	0
U+1F996	T-Rex	nature	1	1	1
U+1F420	Tropical Fish	nature	1	1	0
U+1F422	Turtle	nature	1	1	0
U+2602 U+FE0F	Umbrella	nature	1	1	0
U+1F314	Waxing Gibbous Moon	nature	1	1	0
U+1F940	Wilted Flower	nature	1	1	1
U+1F47C U+1F3FD	Baby Angel: Medium Skin Tone	None	1	1	0
U+1F448 U+1F3FB	Backhand Index Pointing Left: Light Skin Tone	None	1	1	0
U+1F448 U+1F3FC	Backhand Index Pointing Left: Medium-Light Skin Tone	None	1	1	1
U+1F466 U+1F3FB	Boy: Light Skin Tone	None	1	1	0
U+1F919 U+1F3FB	Call Me Hand: Light Skin Tone	None	1	1	1
U+1F919 U+1F3FC	Call Me Hand: Medium-Light Skin Tone	None	1	1	0
U+1F91E U+1F3FE	Crossed Fingers: Medium-Dark Skin Tone	None	1	1	0
U+1F95A	Egg	None	1	1	0
U+2640 U+FE0F	Female Sign	None	1	1	0
U+1F4AA U+1F3FC	Flexed Biceps: Medium-Light Skin Tone	None	1	1	0
U+1F945	Goal Net	None	1	1	0
U+1F647 U+1F3FB U+200D U+2642 U+FE0F	Man Bowing: Light Skin Tone	None	1	1	1
U+1F57A U+1F3FE	Man Dancing: Medium-Dark Skin Tone	None	1	1	0

Codepoint	Emoji Name	Emojipedia Category	Appearances		
			Total	Tweet	Solo
U+1F57A U+1F3FC	Man Dancing: Medium-Light Skin Tone	None	1	1	0
U+1F926 U+1F3FB U+200D U+2642 U+FE0F	Man Facepalming: Light Skin Tone	None	1	1	1
U+1F926 U+1F3FD U+200D U+2642 U+FE0F	Man Facepalming: Medium Skin Tone	None	1	1	1
U+1F926 U+1F3FE U+200D U+2642 U+FE0F	Man Facepalming: Medium-Dark Skin Tone	None	1	1	1
U+1F9DA U+1F3FC U+200D U+2642 U+FE0F	Man Fairy: Medium-Light Skin Tone	None	1	1	1
U+1F645 U+1F3FF U+200D U+2642 U+FE0F	Man Gesturing No: Dark Skin Tone	None	1	1	1
U+1F645 U+1F3FD U+200D U+2642 U+FE0F	Man Gesturing No: Medium Skin Tone	None	1	1	0
U+1F646 U+1F3FF U+200D U+2642 U+FE0F	Man Gesturing OK: Dark Skin Tone	None	1	1	1
U+1F646 U+1F3FB U+200D U+2642 U+FE0F	Man Gesturing OK: Light Skin Tone	None	1	1	1
U+1F646 U+1F3FD U+200D U+2642 U+FE0F	Man Gesturing OK: Medium Skin Tone	None	1	1	0
U+1F486 U+1F3FD U+200D U+2642 U+FE0F	Man Getting Massage: Medium Skin Tone	None	1	1	0
U+1F937 U+1F3FD U+200D U+2642 U+FE0F	Man Shrugging: Medium Skin Tone	None	1	1	0
U+1F937 U+1F3FE U+200D U+2642 U+FE0F	Man Shrugging: Medium-Dark Skin Tone	None	1	1	1
U+1F9DC U+1F3FB U+200D U+2640 U+FE0F	Mermaid: Light Skin Tone	None	1	1	0
U+1F595 U+1F3FB	Middle Finger: Light Skin Tone	None	1	1	0
U+1F595 U+1F3FD	Middle Finger: Medium Skin Tone	None	1	1	0
U+1F595 U+1F3FE	Middle Finger: Medium-Dark Skin Tone	None	1	1	1
U+1F485 U+1F3FB	Nail Polish: Light Skin Tone	None	1	1	0
U+1F485 U+1F3FD	Nail Polish: Medium Skin Tone	None	1	1	0
U+1F44C U+1F3FD	OK Hand: Medium Skin Tone	None	1	1	1
U+1F44C U+1F3FC	OK Hand: Medium-Light Skin Tone	None	1	1	0
U+1F474 U+1F3FB	Old Man: Light Skin Tone	None	1	1	1
U+1F474 U+1F3FC	Old Man: Medium-Light Skin Tone	None	1	1	0

Codepoint	Emoji Name	Emojipedia Category	Appearances		
			Total	Tweet	Solo
U+1F9D3 U+1F3FC	Older Adult: Medium-Light Skin Tone	None	1	1	1
U+1F450 U+1F3FC	Open Hands: Medium-Light Skin Tone	None	1	1	0
U+1F932 U+1F3FE	Palms Up Together: Medium-Dark Skin Tone	None	1	1	1
U+1F64D U+1F3FC	Person Frowning: Medium-Light Skin Tone	None	1	1	1
U+1F646 U+1F3FB	Person Gesturing OK: Light Skin Tone	None	1	1	1
U+1F3C3 U+1F3FD	Person Running: Medium Skin Tone	None	1	1	1
U+1F937 U+1F3FD	Person Shrugging: Medium Skin Tone	None	1	1	0
U+1F481 U+1F3FE	Person Tipping Hand: Medium-Dark Skin Tone	None	1	1	0
U+1F481 U+1F3FC	Person Tipping Hand: Medium-Light Skin Tone	None	1	1	0
U+1F64C U+1F3FC	Raising Hands: Medium-Light Skin Tone	None	1	1	1
U+1F918 U+1F3FE	Sign of the Horns: Medium-Dark Skin Tone	None	1	1	1
U+1F699	Sport Utility Vehicle	None	1	1	0
U+1F44E U+1F3FB	Thumbs Down: Light Skin Tone	None	1	1	1
U+1F44E U+1F3FE	Thumbs Down: Medium-Dark Skin Tone	None	1	1	1
U+1F44D U+1F3FC	Thumbs Up: Medium-Light Skin Tone	None	1	1	0
U+1F647 U+1F3FB U+200D U+2640 U+FE0F	Woman Bowing: Light Skin Tone	None	1	1	1
U+1F647 U+1F3FD U+200D U+2640 U+FE0F	Woman Bowing: Medium Skin Tone	None	1	1	0
U+1F938 U+1F3FB U+200D U+2640 U+FE0F	Woman Cartwheeling: Light Skin Tone	None	1	1	0
U+1F483 U+1F3FE	Woman Dancing: Medium-Dark Skin Tone	None	1	1	1
U+1F483 U+1F3FC	Woman Dancing: Medium-Light Skin Tone	None	1	1	0
U+1F9DD U+1F3FE U+200D U+2640 U+FE0F	Woman Elf: Medium-Dark Skin Tone	None	1	1	1
U+1F926 U+1F3FF U+200D U+2640 U+FE0F	Woman Facepalming: Dark Skin Tone	None	1	1	1
U+1F926 U+1F3FD U+200D U+2640 U+FE0F	Woman Facepalming: Medium Skin Tone	None	1	1	0
U+1F64D U+1F3FB U+200D U+2640 U+FE0F	Woman Frowning: Light Skin Tone	None	1	1	1

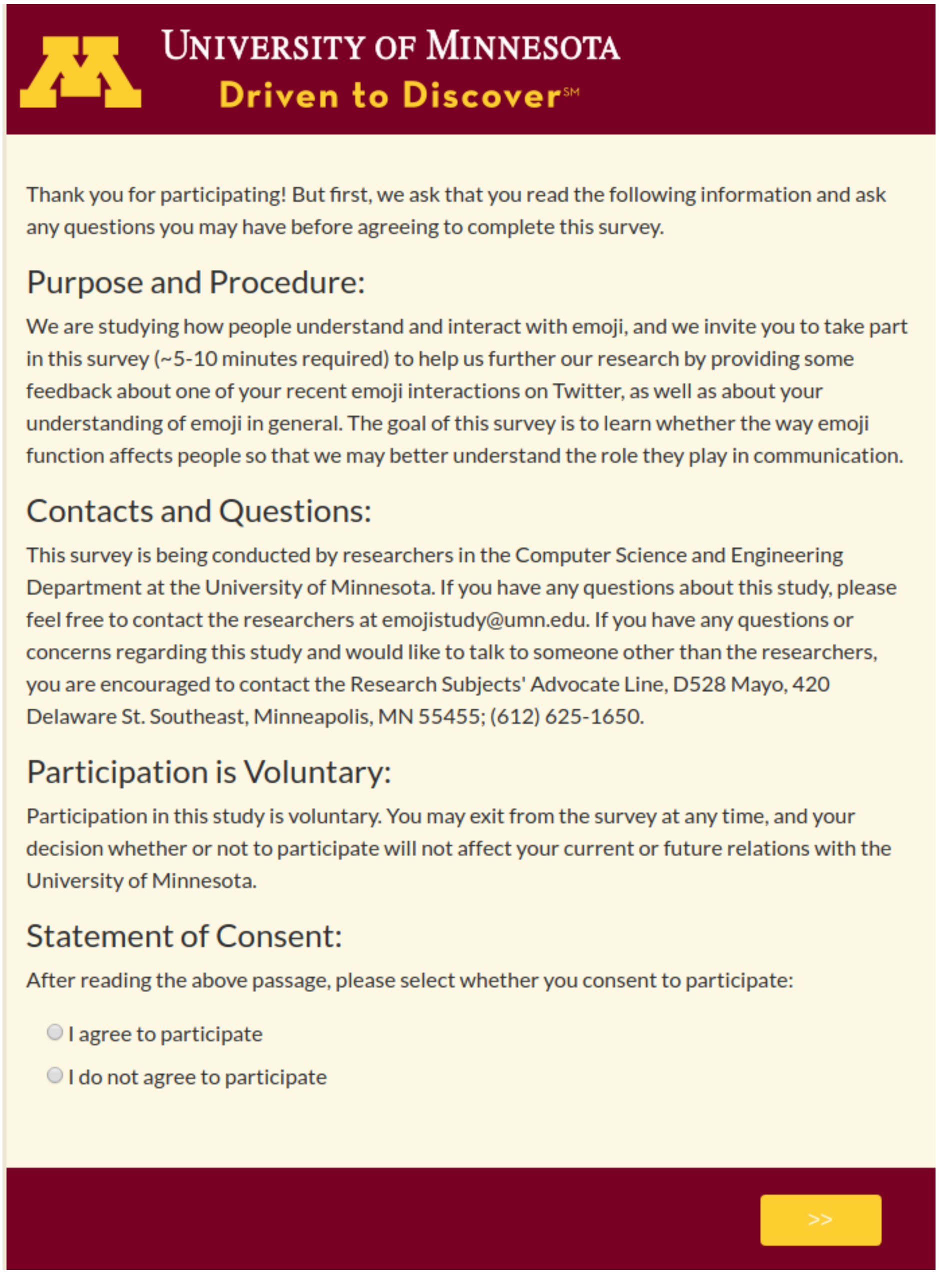
Codepoint	Emoji Name	Emojipedia Category	Appearances		
			Total	Tweet	Solo
U+1F64D U+1F3FD U+200D U+2640 U+FE0F	Woman Frowning: Medium Skin Tone	None	1	1	1
U+1F645 U+1F3FB U+200D U+2640 U+FE0F	Woman Gesturing No: Light Skin Tone	None	1	1	1
U+1F645 U+1F3FD U+200D U+2640 U+FE0F	Woman Gesturing No: Medium Skin Tone	None	1	1	1
U+1F645 U+1F3FE U+200D U+2640 U+FE0F	Woman Gesturing No: Medium-Dark Skin Tone	None	1	1	1
U+1F646 U+1F3FD U+200D U+2640 U+FE0F	Woman Gesturing OK: Medium Skin Tone	None	1	1	0
U+1F486 U+1F3FB U+200D U+2640 U+FE0F	Woman Getting Massage: Light Skin Tone	None	1	1	0
U+1F486 U+1F3FD U+200D U+2640 U+FE0F	Woman Getting Massage: Medium Skin Tone	None	1	1	1
U+1F9D8 U+1F3FB U+200D U+2640 U+FE0F	Woman in Lotus Position: Light Skin Tone	None	1	1	0
U+1F64B U+1F3FB U+200D U+2640 U+FE0F	Woman Raising Hand: Light Skin Tone	None	1	1	0
U+1F64B U+1F3FD U+200D U+2640 U+FE0F	Woman Raising Hand: Medium Skin Tone	None	1	1	1
U+1F937 U+1F3FF U+200D U+2640 U+FE0F	Woman Shrugging: Dark Skin Tone	None	1	1	1
U+1F937 U+1F3FE U+200D U+2640 U+FE0F	Woman Shrugging: Medium-Dark Skin Tone	None	1	1	1
U+1F937 U+1F3FC U+200D U+2640 U+FE0F	Woman Shrugging: Medium-Light Skin Tone	None	1	1	0
U+1F469 U+1F3FF U+200D U+1F3EB	Woman Teacher: Dark Skin Tone	None	1	1	1
U+1F481 U+1F3FD U+200D U+2640 U+FE0F	Woman Tipping Hand: Medium Skin Tone	None	1	1	0
U+1F481 U+1F3FC U+200D U+2640 U+FE0F	Woman Tipping Hand: Medium-Light Skin Tone	None	1	1	0
U+1F9DB U+1F3FB U+200D U+2640 U+FE0F	Woman Vampire: Light Skin Tone	None	1	1	0
U+1F4CA	Bar Chart	objects	1	1	1

Codepoint	Emoji Name	Emojipedia Category	Appearances		
			Total	Tweet	Solo
U+1F6CF U+FE0F	Bed	objects	1	1	1
U+1F4A3	Bomb	objects	1	1	0
U+1F4DA	Books	objects	1	1	1
U+1F39B U+FE0F	Control Knobs	objects	1	1	0
U+1F4B3	Credit Card	objects	1	1	0
U+1F5A5 U+FE0F	Desktop Computer	objects	1	1	0
U+1F4B5	Dollar Banknote	objects	1	1	0
U+1F4C0	DVD	objects	1	1	0
U+1F4FD U+FE0F	Film Projector	objects	1	1	1
U+231B	Hourglass Done	objects	1	1	0
U+23F3	Hourglass Not Done	objects	1	1	0
U+1F38E	Japanese Dolls	objects	1	1	0
U+1F52A	Kitchen Knife	objects	1	1	0
U+1F4A1	Light Bulb	objects	1	1	1
U+1F4DD	Memo	objects	1	1	0
U+1F4F0	Newspaper	objects	1	1	0
U+1F5DD U+FE0F	Old Key	objects	1	1	0
U+1F4D6	Open Book	objects	1	1	1
U+1F4BF	Optical Disk	objects	1	1	0
U+1F5A8 U+FE0F	Printer	objects	1	1	0
U+1F6BF	Shower	objects	1	1	1
U+23F1 U+FE0F	Stopwatch	objects	1	1	0
U+1F4C6	Tear-Off Calendar	objects	1	1	1
U+260E U+FE0F	Telephone	objects	1	1	0
U+1F4DE	Telephone Receiver	objects	1	1	1
U+23F2 U+FE0F	Timer Clock	objects	1	1	0
U+1F6BD	Toilet	objects	1	1	1
U+1F5D1 U+FE0F	Wastebasket	objects	1	1	1
U+231A	Watch	objects	1	1	0
U+1F381	Wrapped Gift	objects	1	1	0
U+1F47D	Alien	people	1	1	0
U+1F476	Baby	people	1	1	1
U+1F448	Backhand Index Pointing Left	people	1	1	0
U+1F449	Backhand Index Pointing Right	people	1	1	0
U+1F446	Backhand Index Pointing Up	people	1	1	0
U+1F471 U+200D U+2640 U+FE0F	Blond-Haired Woman	people	1	1	1
U+1F9E0	Brain	people	1	1	0
U+1F4BC	Briefcase	people	1	1	0
U+1F921	Clown Face	people	1	1	0
U+1F491	Couple With Heart	people	1	1	1
U+1F920	Cowboy Hat Face	people	1	1	0
U+1F442	Ear	people	1	1	1
U+1F92E	Face Vomiting	people	1	1	0
U+1F915	Face With Head-Bandage	people	1	1	1
U+1F92C	Face With Symbols on Mouth	people	1	1	1
U+1F46A	Family	people	1	1	0
U+1F469 U+200D U+1F469 U+200D U+1F467 U+200D U+1F467	Family: Woman, Woman, Girl, Girl	people	1	1	0

Codepoint	Emoji Name	Emojipedia Category	Appearances		
			Total	Tweet	Solo
U+1F47B	Ghost	people	1	1	1
U+1F467	Girl	people	1	1	0
U+1F453	Glasses	people	1	1	0
U+1F638	Grinning Cat Face With Smiling Eyes	people	1	1	1
U+1F600	Grinning Face	people	1	1	0
U+1F45C	Handbag	people	1	1	0
U+1F460	High-Heeled Shoe	people	1	1	0
U+1F617	Kissing Face	people	1	1	1
U+1F484	Lipstick	people	1	1	0
U+1F468	Man	people	1	1	0
U+1F9DA U+200D U+2642 U+FE0F	Man Fairy	people	1	1	0
U+1F468 U+200D U+2695 U+FE0F	Man Health Worker	people	1	1	0
U+1F6B6 U+200D U+2642 U+FE0F	Man Walking	people	1	1	1
U+1F64D	Person Frowning	people	1	1	1
U+1F487	Person Getting Haircut	people	1	1	1
U+1F3C3	Person Running	people	1	1	1
U+1F930	Pregnant Woman	people	1	1	0
U+1F91A	Raised Back of Hand	people	1	1	1
U+1F480	Skull	people	1	1	0
U+1F63B	Smiling Cat Face With Heart-Eyes	people	1	1	1
U+1F61D	Squinting Face With Tongue	people	1	1	1
U+1F44D	Thumbs Up	people	1	1	1
U+1F9DA U+200D U+2640 U+FE0F	Woman Fairy	people	1	1	0
U+1F486 U+200D U+2640 U+FE0F	Woman Getting Massage	people	1	1	1
U+1F469 U+200D U+1F52C	Woman Scientist	people	1	1	1
U+1F9DF U+200D U+2640 U+FE0F	Woman Zombie	people	1	1	0
U+269B U+FE0F	Atom Symbol	symbols	1	1	0
U+1F519	Back Arrow	symbols	1	1	0
U+25FC U+FE0F	Black Medium Square	symbols	1	1	0
U+25AA U+FE0F	Black Small Square	symbols	1	1	0
U+1F535	Blue Circle	symbols	1	1	0
U+1F506	Bright Button	symbols	1	1	1
U+1F3A6	Cinema	symbols	1	1	0
U+1F300	Cyclone	symbols	1	1	0
U+2666 U+FE0F	Diamond Suit	symbols	1	1	0
U+1F51A	End Arrow	symbols	1	1	1
U+2714 U+FE0F	Heavy Check Mark	symbols	1	1	0
U+2796	Heavy Minus Sign	symbols	1	1	1
U+1F4E2	Loudspeaker	symbols	1	1	0
U+1F550	One O’clock	symbols	1	1	0
U+1F55C	One-Thirty	symbols	1	1	0
U+303D U+FE0F	Part Alternation Mark	symbols	1	1	0
U+1F53B	Red Triangle Pointed Down	symbols	1	1	0
U+AE U+FE0F	Registered	symbols	1	1	0

Codepoint	Emoji Name	Emojipedia Category	Appearances		
			Total	Tweet	Solo
U+1F556	Seven O’clock	symbols	1	1	0
U+1F562	Seven-Thirty	symbols	1	1	0
U+2747 U+FE0F	Sparkle	symbols	1	1	1
U+1F4AC	Speech Balloon	symbols	1	1	0
U+2721 U+FE0F	Star of David	symbols	1	1	0
U+1F552	Three O’clock	symbols	1	1	0
U+26A0 U+FE0F	Warning	symbols	1	1	0
U+1F6BE	Water Closet	symbols	1	1	0
U+26AA	White Circle	symbols	1	1	0
U+25AB U+FE0F	White Small Square	symbols	1	1	0
U+1F4A4	Zzz	symbols	1	1	1
U+1F697	Automobile	travel-places	1	1	0
U+1F3D6 U+FE0F	Beach With Umbrella	travel-places	1	1	1
U+1F3F0	Castle	travel-places	1	1	1
U+26EA	Church	travel-places	1	1	0
U+1F692	Fire Engine	travel-places	1	1	1
U+1F301	Foggy	travel-places	1	1	0
U+1F684	High-Speed Train	travel-places	1	1	0
U+1F3E0	House	travel-places	1	1	1
U+1F688	Light Rail	travel-places	1	1	0
U+1F687	Metro	travel-places	1	1	0
U+1F30C	Milky Way	travel-places	1	1	0
U+1F69D	Monorail	travel-places	1	1	0
U+1F3CD U+FE0F	Motorcycle	travel-places	1	1	0
U+1F3DE U+FE0F	National Park	travel-places	1	1	1
U+1F68D	Oncoming Bus	travel-places	1	1	0
U+1F683	Railway Car	travel-places	1	1	0
U+1F680	Rocket	travel-places	1	1	0
U+1F320	Shooting Star	travel-places	1	1	0
U+1F689	Station	travel-places	1	1	0
U+1F305	Sunrise	travel-places	1	1	0
U+1F304	Sunrise Over Mountains	travel-places	1	1	1
U+1F686	Train	travel-places	1	1	0
U+1F68A	Tram	travel-places	1	1	0
U+1F68B	Tram Car	travel-places	1	1	0

Figure A.1: Example Run Through the Survey from Chapter 5



UNIVERSITY OF MINNESOTA
Driven to Discover℠

Thank you for participating! But first, we ask that you read the following information and ask any questions you may have before agreeing to complete this survey.

Purpose and Procedure:

We are studying how people understand and interact with emoji, and we invite you to take part in this survey (~5-10 minutes required) to help us further our research by providing some feedback about one of your recent emoji interactions on Twitter, as well as about your understanding of emoji in general. The goal of this survey is to learn whether the way emoji function affects people so that we may better understand the role they play in communication.

Contacts and Questions:

This survey is being conducted by researchers in the Computer Science and Engineering Department at the University of Minnesota. If you have any questions about this study, please feel free to contact the researchers at emojistudy@umn.edu. If you have any questions or concerns regarding this study and would like to talk to someone other than the researchers, you are encouraged to contact the Research Subjects' Advocate Line, D528 Mayo, 420 Delaware St. Southeast, Minneapolis, MN 55455; (612) 625-1650.

Participation is Voluntary:

Participation in this study is voluntary. You may exit from the survey at any time, and your decision whether or not to participate will not affect your current or future relations with the University of Minnesota.

Statement of Consent:

After reading the above passage, please select whether you consent to participate:

- I agree to participate
- I do not agree to participate

>>

UNIVERSITY OF MINNESOTA

Driven to Discover℠

This survey is tailored specifically to you and your twitter activity. Please enter your Twitter username or "handle" to help us confirm that we have the correct data:

@ handle

Note: Your Twitter username begins with "@" like @UMNEmojiStudy, and you can find it below your display name in your profile information.

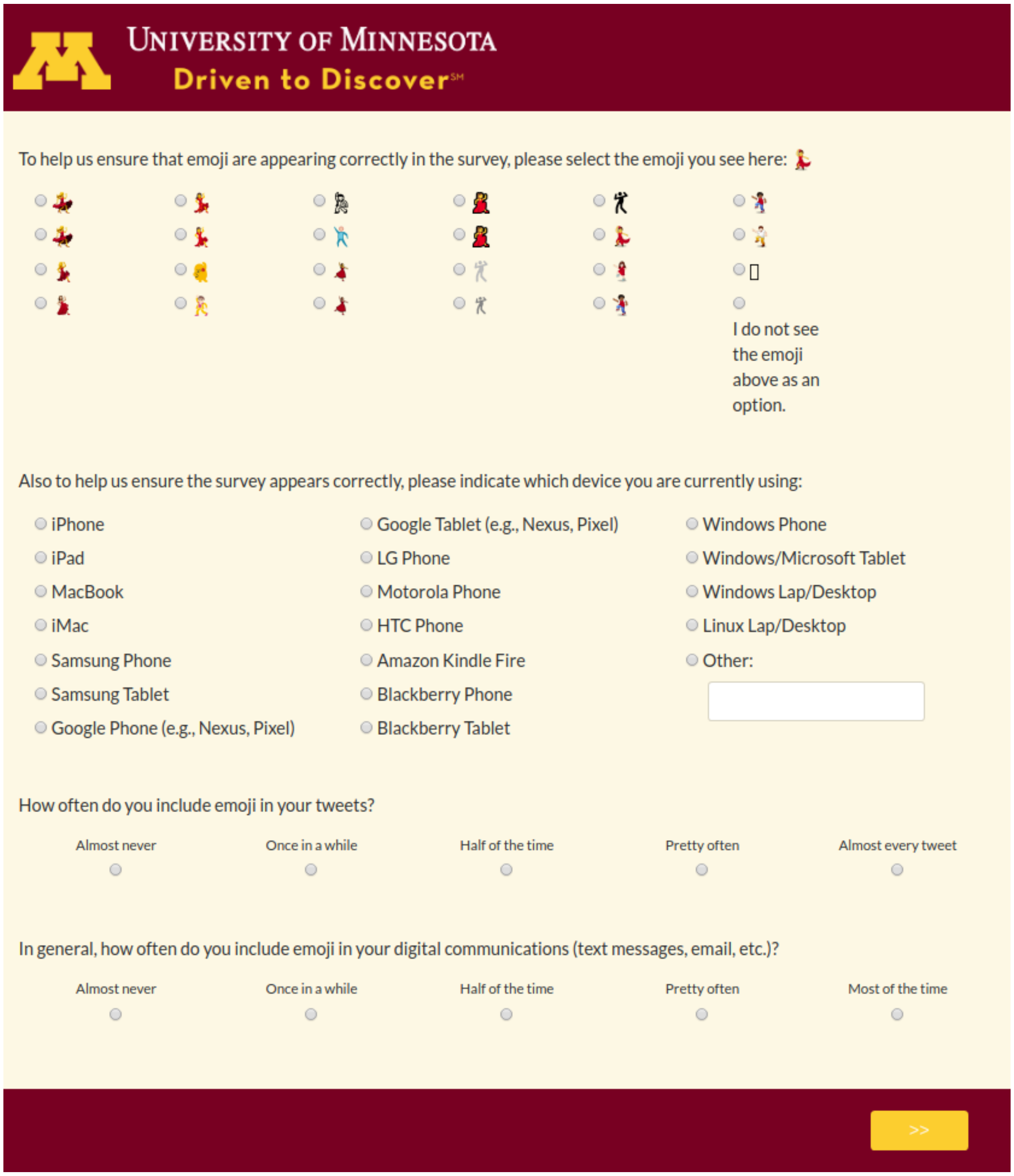
Please select your gender:

- Male
- Female
- Other

Please select your age:

- < 18
- 18-25
- 26-35
- 36-45
- 46-55
- 56+

>>



UNIVERSITY OF MINNESOTA
Driven to Discover℠

To help us ensure that emoji are appearing correctly in the survey, please select the emoji you see here:

I do not see the emoji above as an option.

Also to help us ensure the survey appears correctly, please indicate which device you are currently using:

- iPhone
- iPad
- MacBook
- iMac
- Samsung Phone
- Samsung Tablet
- Google Phone (e.g., Nexus, Pixel)
- Google Tablet (e.g., Nexus, Pixel)
- LG Phone
- Motorola Phone
- HTC Phone
- Amazon Kindle Fire
- Blackberry Phone
- Blackberry Tablet
- Windows Phone
- Windows/Microsoft Tablet
- Windows Lap/Desktop
- Linux Lap/Desktop
- Other:

How often do you include emoji in your tweets?

Almost never	Once in a while	Half of the time	Pretty often	Almost every tweet

In general, how often do you include emoji in your digital communications (text messages, email, etc.)?

Almost never	Once in a while	Half of the time	Pretty often	Most of the time

>>

University of Minnesota
Driven to Discover℠

You recently tweeted:

> Not long ago, we had something in common, But the winds kept blowing and changed us both

Do you think that this tweet will appear exactly this way to everyone who views it?

- Yes
- No

Please explain your answer (optional):

>>

UNIVERSITY OF MINNESOTA
Driven to Discover℠

You recently tweeted:

> Not long ago, we had something in common, But the winds kept blowing and changed us both

How important is the emoji to this tweet? Please select your level of agreement with the following statements:

This tweet needs this emoji to convey what I meant.

Strongly Disagree	Disagree	Not Sure	Agree	Strongly Agree
○	○	○	○	○

The emoji could be removed from this tweet and it would not make a difference.

Strongly Disagree	Disagree	Not Sure	Agree	Strongly Agree
○	○	○	○	○

A different emoji could be substituted for this emoji in this tweet and not change my meaning.

Strongly Disagree	Disagree	Not Sure	Agree	Strongly Agree
○	○	○	○	○

>>

University of Minnesota
Driven to Discover℠

You recently tweeted:

> Not long ago, we had something in common, But the winds kept blowing and changed us both

Did you know that the emoji in your tweet will appear differently to other users on Twitter? For example, your tweet will appear as the following on the associated devices / operating systems:

Not long ago, we had something in common, But the winds kept blowing and changed us both	Apple iOS 11.2, 11.1
Not long ago, we had something in common, But the winds kept blowing and changed us both ▯ ▯	Apple iOS 10.3, 10.2, 10.0, 9.3, 9.1, 9.0, 8, 6/7
Not long ago, we had something in common, But the winds kept blowing and changed us both	Google Android 8.1, 8.0
Not long ago, we had something in common, But the winds kept blowing and changed us both ▯	Google Android 7.1, 7.0, 6.0.1, 5.0, 4.4
Not long ago, we had something in common, But the winds kept blowing and changed us both ▯▯▯	HTC Devices
Not long ago, we had something in common, But the winds kept blowing and changed us both ▯ ▯	LG G5, G4
Not long ago, we had something in common, But the winds kept blowing and changed us both	Microsoft Windows 10
Not long ago, we had something in common, But the winds kept blowing and changed us both ▯▯▯	Microsoft Windows 8, 7, 8 (Not Updated)
Not long ago, we had something in common, But the winds kept blowing and changed us both ▯	Samsung Devices
Not long ago, we had something in common, But the winds kept blowing and changed us both	Twitter Web Client (Twemoji)

- ○ Yes, I knew this.
- ○ No, I did not know this.

>>

University of Minnesota
Driven to Discover℠

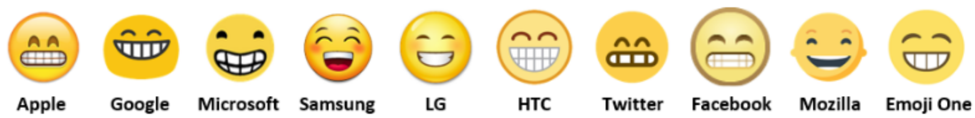
Since you said you did not know, here is a little explanation:

To your device, an emoji is just like any other character (e.g., lower-case 'a', upper-case 'B') and needs to be rendered with a font (e.g., Calibri, Times New Roman). However, for emoji, fonts are unique to device and communication platform vendors. For example, Apple has its own emoji font for iOS/macOS devices (e.g., iPhone, iPad, MacBook), Samsung has its own emoji font for Samsung devices (e.g., Galaxy phones, tablets), etc. Twitter has its own emoji font for when Twitter is viewed in a browser, but users see their own device's emoji when they view Twitter via a device's mobile Twitter app.

All of this means that a given emoji character looks different on different device platforms:

These are all the same emoji!

This is what the "grinning face with smiling eyes" emoji looks like on devices for each of these platforms:



So when you use an emoji, you see your device's rendition of the emoji. But when your followers view that emoji, they will see their device's rendition of the emoji. If your devices have the same emoji font, then you will both see the same rendition of the emoji. But if your devices have different emoji fonts, then you will both see different renditions of the emoji.

How would you describe your reaction to finding out that this is how emoji function? (Optional)

If you had to summarize your reaction in one or two words, what would it be?

>>

University of Minnesota
Driven to Discover

Your tweet (what you see):

Not long ago, we had something in common, But the winds kept blowing and changed us both

What your followers see:

Not long ago, we had something in common, But the winds kept blowing and changed us both	Apple iOS 11.2, 11.1
Not long ago, we had something in common, But the winds kept blowing and changed us both ▯ ▯	Apple iOS 10.3, 10.2, 10.0, 9.3, 9.1, 9.0, 8, 6/7
Not long ago, we had something in common, But the winds kept blowing and changed us both	Google Android 8.1, 8.0
Not long ago, we had something in common, But the winds kept blowing and changed us both ▯	Google Android 7.1, 7.0, 6.0.1, 5.0, 4.4
Not long ago, we had something in common, But the winds kept blowing and changed us both ▯▯▯	HTC Devices
Not long ago, we had something in common, But the winds kept blowing and changed us both ▯ ▯	LG G5, G4
Not long ago, we had something in common, But the winds kept blowing and changed us both	Microsoft Windows 10
Not long ago, we had something in common, But the winds kept blowing and changed us both ▯▯▯	Microsoft Windows 8, 7, 8 (Not Updated)
Not long ago, we had something in common, But the winds kept blowing and changed us both ▯	Samsung Devices
Not long ago, we had something in common, But the winds kept blowing and changed us both	Twitter Web Client (Twemoji)

Do you think your followers' versions of the tweet convey the same message you intended to send with your tweet?

- **Yes**, I think my followers' versions convey the same message.
- I think **some** of my followers' versions convey the same message, some do not.
- **No**, I think my followers' versions do not convey the same message.

Please explain your answer (optional):

Do you think your followers will interpret your tweet the same way you do?

- **Yes**, I think my followers will interpret my tweet the same way.
- I think **some** of my followers will interpret my tweet the same way, some will not.
- **No**, I think my followers will not interpret my tweet the same way.

Please explain your answer (optional):

Do you think your followers will interpret your tweet the same way you do?

- **Yes**, I think my followers will interpret my tweet the same way.
- I think **some** of my followers will interpret my tweet the same way, some will not.
- **No**, I think my followers will not interpret my tweet the same way.

Please explain your answer (optional):

If you had known that this is how your tweet would look to your audience, would you have sent it as is?

- Yes
- No

Please explain your answer (optional):

How would you edit your tweet knowing this is how it looks to your audience?

- I would not edit my tweet.
- I would edit the text in my tweet.
- I would add another emoji to my tweet.
- I would replace the emoji with another in my tweet.
- I would remove the emoji from my tweet.
- Other:

>>

University of Minnesota
Driven to Discover℠

Now that you are aware that emoji appear differently when you communicate across platforms, please describe your general impression of this:

In general, do you think this may have any effect on your Twitter communication?

- Yes
- No

Please explain your answer (optional):

This is also the way emoji function in communication across platforms outside of Twitter, like in text messaging for example. Considering this, do you think this may have any effect on your direct communication outside of Twitter (e.g., when you directly text a friend)?

- Yes
- No

Please explain your answer (optional):

>>

UNIVERSITY OF MINNESOTA
Driven to Discover℠

When you tweet, who do you feel like you're typically targeting? How would you describe your Twitter following (i.e., those that follow you on Twitter)?

Does your Twitter following contain... (please check all that apply)

- ☐ Friends
- ☐ Family
- ☐ Professional Connections
- ☐ Online-only Connections
- ☐ Strangers
- ☐ Other:

Please indicate all devices that you use on a regular basis:

- ☐ iPhone
- ☐ iPad
- ☐ MacBook
- ☐ iMac
- ☐ Samsung Phone
- ☐ Samsung Tablet
- ☐ Google Phone (e.g., Nexus, Pixel)
- ☐ Google Tablet (e.g., Nexus, Pixel)
- ☐ LG Phone
- ☐ Motorola Phone
- ☐ HTC Phone
- ☐ Amazon Kindle Fire
- ☐ Blackberry Phone
- ☐ Blackberry Tablet
- ☐ Windows Phone
- ☐ Windows/Microsoft Tablet
- ☐ Windows Lap/Desktop
- ☐ Linux Lap/Desktop
- ☐ Other:

Please indicate which of the following applications you use:

- ☐ Text Messages
- ☐ Google Hangouts
- ☐ Gmail
- ☐ Email (not Gmail)
- ☐ Facebook
- ☐ Facebook Messenger
- ☐ Instagram
- ☐ Snapchat
- ☐ Slack
- ☐ WhatsApp

>>

UNIVERSITY OF MINNESOTA
Driven to Discover℠

We may continue this research and we would like to know: Are you open to us contacting you again for future participation?

- Yes, you may contact me again in the future.
- No, please do not contact me again.

Is there anything you'd like to share with us before submitting your survey? Any comments, feedback, suggestions? (optional)

>>

UNIVERSITY OF MINNESOTA
Driven to Discover℠

Your survey has been submitted. Thank you so much for your time and participation. Happy tweeting!