Proceedings of the Tenth International AAAI Conference on
Web and Social Media (ICWSM 2016)

"Blissfully Happy" or "Ready to Fight": Varying Interpretations of Emoji

Hannah Miller, Jacob Thebault-Spieker, Shuo Chang, Isaac Johnson,
Loren Terveen, Brent Hecht
GroupLens Research, University of Minnesota
Minneapolis, MN 55455, USA
{hmiller, thebault, schang, ijohnson, terveen, bhecht}@cs.umn.edu

Abstract

Emoji are commonly used in modern text communication. However, as graphics 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. We explore whether emoji renderings or differences across platforms give rise to diverse interpretations of emoji. Through an online survey, we solicit people's interpretations of a sample of the most popular emoji characters, each rendered for multiple platforms. Both in terms of sentiment and semantics, we analyze the variance in interpretation of the emoji, quantifying which emoji are most (and least) likely to be misinterpreted. In cases in which participants rated the same emoji rendering, they disagreed on whether the sentiment was positive, neutral, or negative 25% of the time. When considering renderings across platforms, these disagreements only increase. Overall, we find significant potential for miscommunication, both for individual emoji renderings and for different emoji renderings across platforms.

Introduction

Emoji are "picture characters" or pictographs that are popular in text-based communication. They are commonly used in smartphone texting, social media sharing (e.g., nearly half of all text on Instagram contains emoji (Dimson 2015)), advertising (e.g., Chevy's press release written entirely in emoji #ChevyGoesEmoji[1]), and more. Oxford Dictionaries declared the emoji 😂 or "face with tears of joy" to be the 2015 "word of the year." As this is the first time that they had selected an emoji, they noted that "emoji have come to embody a core aspect of living in a digital world that is visually driven, emotionally expressive, and obsessively immediate."[2]

Most commonly-used emoji are encoded in the Unicode standard for indexing characters[3]. There are currently 1,282 emoji in the Unicode standard, and for each of these, the Unicode Consortium provides a code and name (e.g., U+1F600 for "grinning face") but not the actual graphic. This is the same as is the case for Unicode text characters: for example, the Unicode character U+0041 indexes the Latin capital letter 'A', but it does not indicate specifically how the 'A' should look. Instead, a font renders the Unicode characters a particular way: the appearance of this text that you are reading is dictated by the Times New Roman font.

Similarly, individual platform vendors such as Apple and Google create their own *rendering* for each emoji character they support. This means that the "Grinning Face" emoji character has a different appearance when viewed on an Apple device (e.g., an iPhone) than on a Google device (e.g., a Nexus phone). This is just one example of two different platform renderings; there are *many* platforms that each have their own unique set of emoji renderings. Emojipedia—a website serving as an "encyclopedia for emoji"—lists 17 such platforms[4], which means that there may be at least 17 different renderings for a given Unicode emoji 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 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.

[1] http://www.chevrolet.com/crack-the-emoji-code.html
[2] http://time.com/4114886/oxford-word-of-the-year-2015-emoji/
[3] http://unicode.org/emoji/charts/full-emoji-list.html
[4] http://emojipedia.org/

We contextualize our analysis in Herbert Clark's psycholinguistic theory of language use (Clark 1996). In social psychology, 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. In the context of emoji, a speaker is sending emoji to an addressee through a mobile or desktop platform. Likewise, the addressee is receiving the emoji via a mobile or desktop platform. In this exchange, misconstrual can arise from differing interpretations derived from either (1) the same rendering, if they each see the same rendering or (2) different renderings, if they each see a different rendering.

We explore the potential for misconstrual when using emoji in communication by evaluating variation in emoji interpretation. Using an online survey, we solicit people's interpretations of a sample of the most popular emoji Unicode characters. In order to analyze how emoji interpretations vary for renderings across platforms, the survey included renderings of each emoji from five major mobile platforms: Apple, Google, Microsoft, Samsung, and LG. We identify the variance of interpretation in terms of sentiment (i.e., how positive is this emoji?) and semantics (i.e., what does this emoji mean?).

We find that only 4.5% of emoji symbols we 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 increase. For U+1F601 ("grinning face with smiling eyes" according to the Unicode Standard), participants described the Google rendering  as "blissfully happy" while the exact same Unicode character, but rendered for Apple , was described as "ready to fight." We conclude that emoji usage may be ripe for misconstrued communication and provide implications for design to manage the likelihood of misinterpretation when using emoji.

Related Work

We begin this section with a discussion of the role of emoticons (e.g., :-) or 'smiley face' – a precursor to emoji) in the interpretation of text-based communication and how emoji relate to emoticons. We then discuss what we know about the consistency of interpretation for emoticons and emoji.

Emoticons

Emoticons, or "typographic symbols that appear sideways as resembling facial expressions," (Walther and D'Addario 2001) such as **:)**, have been in use in text-based communication since at least the early 1980s, with numerous studies documenting their prevalence in SMS texts (Tossell et al. 2012), blogs (Huffaker and Calvert 2006), and, more recently, Twitter (Park et al. 2013). Much research has focused on the role that emoticons can play in complementing traditional text-based computer-mediated communication (CMC). Notably, Walther and D'Addario (2001) found that while the emotional valence of text (e.g., "I am happy") tends to be more important than any accompanying emoticons with respect to interpretation, a negative emoticon (e.g., **:(** "frowny face") can significantly change the interpretation of the message. Lo (2008) 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. Derks, Fischer, and Bos (2008) concluded in a survey of emotion in CMC that emoticons largely function as non-verbal cues do in face-to-face communication. Going beyond interpretation of individual messages, Liebman and Gergle (2016) demonstrated that emoticons (along with punctuation) are important in interpersonal relationship development over text-based communication. Together, this work emphasizes the importance of emoticons in text-based communication.

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 (Davis and Edberg 2015). They have become quite popular since then, with, for example, over 2% of tweets (Novak et al. 2015) and nearly half of text on Instagram (Dimson 2015) containing emoji. Emoji are often described as a successor to emoticons (e.g., Novak et al. 2015), and Pavalanathan and Eisenstein (2016) found that while emoticons are decreasing in popularity on Twitter, emoji are increasing in popularity and seem to be replacing, not complementing, emoticons.

While the large body of work on the role of emoticons in text-based communication has largely not been replicated for emoji, early work indicates that emoji do fulfill much the same role. Kelly and Watts (2015) 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.

Consistency of Emoticon and Emoji Interpretation

Whereas the display of emoji is platform-dependent, emoticons, as text, are displayed relatively consistently. Walther and D'Addario (2001) found high agreement across their participants (226 mostly male students) around sentiment interpretations of the three emoticons that they studied, :-) and :-(and ;-). In research on using emoticons in sentiment analysis, Davidov, Tsur, and Rappoport (2010) 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.

Less is known about the consistency of emoji interpretation. Researchers such as Liu, Li, and Guo (2012) and Novak et al. (2015) have developed classifiers of emoji sentiment by labeling emoji with the sentiment of the surrounding text. While this has proven largely effective, both papers mentioned instances of emoji being associated with different, and occasionally opposite, sentiment labels. We know of no work, however, that has investigated how the interpretation of emoji varies. We seek to address this gap in the literature and also to understand how the platform-dependence of emoji implementation might further complicate interpretation.

Research Questions

As noted above, each platform has its own unique rendering of emoji Unicode characters (e.g., see Figure 1). Communication can take place *within* platform or *across* platform. If the sender and the receiver are both using the same platform, then they are communicating *within platform* and they see the same emoji rendering. If they are using different platforms, then they are communicating *across platform* and see different renderings of emoji. We break down the goal of learning whether people interpret emoji the same way or not into two research questions based on within- and across-platform communication:

- **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 Platform)**: 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?

We examine interpretation agreement and disagreement along two dimensions: *sentiment* and *semantics*. Sentiment analysis involves "classifying the polarity of a given text."[5] For our purposes, this means determining whether the expression of a given emoji is positive, negative, or neutral. In our context, semantics refers to what people think a given emoji means. For each of our research questions, we explore how people's interpretations manifest **(a)** sentiment and **(b)** semantic differences.

[5] https://en.wikipedia.org/wiki/Sentiment_analysis

Survey

We created an online survey to solicit people's interpretations of a sample of emoji Unicode characters, each rendered for multiple platforms.

Emoji Unicode Character Sample

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

We restricted our sampling to anthropomorphic emoji, or those that represent faces or people, because (1) they are very common and (2) we 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 our Twitter dataset, and SwiftKey (2015) reports that faces or smileys comprise 59% of emoji characters typed with their smartphone keyboard app. We selected the top 25 most popular anthropomorphic emoji Unicode characters for our sample.

Platform Selection

To investigate how people interpret renderings from different platforms, we solicited people's interpretations of multiple platform renderings of each emoji Unicode character in our sample, focusing on smartphone platforms. Using comScore reports from 2015[6], we picked the top three smartphone platforms: Android, Apple, and Microsoft. Since Android is fragmented by manufacturer, we selected Google's rendering, as well as the renderings of the top two Android hardware manufacturers: Samsung and LG[7]. We used renderings for these five platforms for every Unicode character in our study. To collect the graphics of the emoji to use in our survey, we used data from Emojipedia[8].

Survey Design

With 5 platform renderings of 25 emoji Unicode characters, we gathered survey results for 125 total emoji renderings. We employed a purely random between-subjects design, and each participant received a random sample of 15 emoji renderings to interpret from the 125 total. We aimed

[6] https://www.comscore.com/Insights/Market-Rankings/comScore-Reports-July-2015-US-Smartphone-Subscriber-Market-Share

[7] 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.

[8] http://emojipedia.org/

to collect approximately 40 interpretations per emoji rendering. Thus for a total of 5000 interpretations, and 15 interpretations per participant, we 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] ______________________

Questions one, two, and four elicited text responses and were focused on semantic interpretations of emoji. Question three elicited a numeric sentiment judgment, mirroring the -5 to 5 sentiment scale used in Taboada et al. (2011).

In addition to the survey pages for the emoji in our sample, we created the same page for Apple's heart emoji (❤, Unicode U+2764). We 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 us to control for quality of responses by assessing intra-rater agreement on each participant's two ratings of the heart emoji. We also assessed the variance of participants' overall ratings of the heart emoji, and find that our participants are very consistent in their sentiment evaluation: they vary, on average, by 0.54 (out of 10) sentiment points.

Participants

We recruited our survey participants via Amazon Mechanical Turk. We 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 our survey, we 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), we compensated participants $1.20 for completing the survey.

Our 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, we calculated intra-rater reliability to ensure consistency within each participant's ratings. We 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. We considered these participants to be consistent in their ratings and excluded the remaining 26 participant responses from our dataset. To identify any low-quality participant responses that were not reflected through sentiment rating inconsistency, we 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, we retained the data of 304 participants for our 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".

Data for Analysis

With 304 participants each completing 15 emoji interpretations, we 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 our analysis, we discovered an error in our emoji sample. We cross-checked back with Emojipedia, the site from which we downloaded our emoji images, and discovered that some of the images in our set (automatically labelled by Unicode and platform at the time of download) had been incorrectly labeled at the time of download. We accordingly examined and reorganized our survey data to ensure that we were associating participants' interpretations with the correct emoji rendering. We ended up with incomplete data for 3 of the 25 Unicode emoji characters we sampled, so we excluded them from our analysis (U+1F614 "pensive face," U+1F633 "flushed face," and U+1F604 "smiling face with open mouth and smiling eyes").

Analyses and Results

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

Sentiment Analysis

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

Methods

For each emoji rendering, we 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, we computed the pairwise differences (i.e. distances) of these sentiment scores. 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 variation in interpretation for specific emoji renderings (RQ1), we 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), we performed a similar calculation, but focused on differences in rated sentiment across different platform renderings of the same emoji Unicode character. For a given Unicode character (e.g., "face with tears of joy"), and a pair of platforms (e.g., Apple and LG), we computed all pairwise distances between the two sets of sentiment ratings, and then took the average (e.g. an Apple-LG average sentiment distance). We did this for all pairs of platforms, and ended up with platform-pair average sentiment distances (e.g. one for Apple-LG, one for Apple-Microsoft, one for LG-Microsoft, etc.). We then computed the grand-mean (mean of these average sentiment distances), as the *across-platform sentiment misconstrual score*.

Results

RQ1 (Within Platform) for Sentiment

To understand the extent to which interpretation of the sentiment of each emoji rendering varies, we ranked each rendering based on the within-platform sentiment misconstrual score in descending order for each platform. We present the top three and bottom three of this ranking in Table 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, our participants differed by 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 gave it a 1 or -1.

	Most/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)

Table 1. Top-3 and bottom-3 most different in terms of sentiment. Higher values indicate greater response variation.

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 emoji (even within a single platform) is 2 or more. On the other hand, only five renderings (4.5%) have a misconstrual score of 1 or less.

We also report the average sentiment misconstrual score across all Unicode characters for each platform in Table 1. Apple has the highest average within-platform sentiment misconstrual (1.96); Google has the lowest (1.79).

Overall, we 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 smartphone platform), there is still plenty of sentiment misconstrual. Indeed, if we select two participants who have rated the exact same rendering, in 25% of those cases, they did not agree on whether the sentiment was positive, neutral, or negative. This reflects the most straightforward form of *within-platform* communication, and our results suggest that, even in this case, there are clear opportunities for misconstrued communication.

RQ2 (Across Platform) for Sentiment

We now explore variance in sentiment for renderings across platforms. In Figure 1, we 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 Unicode characters (each set of five renderings are shown along the x-axis in Figure 1). We find that approximately 41% (9 of 22) of the Unicode characters have a range wider than one sentiment unit, suggesting that at least one platform's rendering of these Unicode characters is different

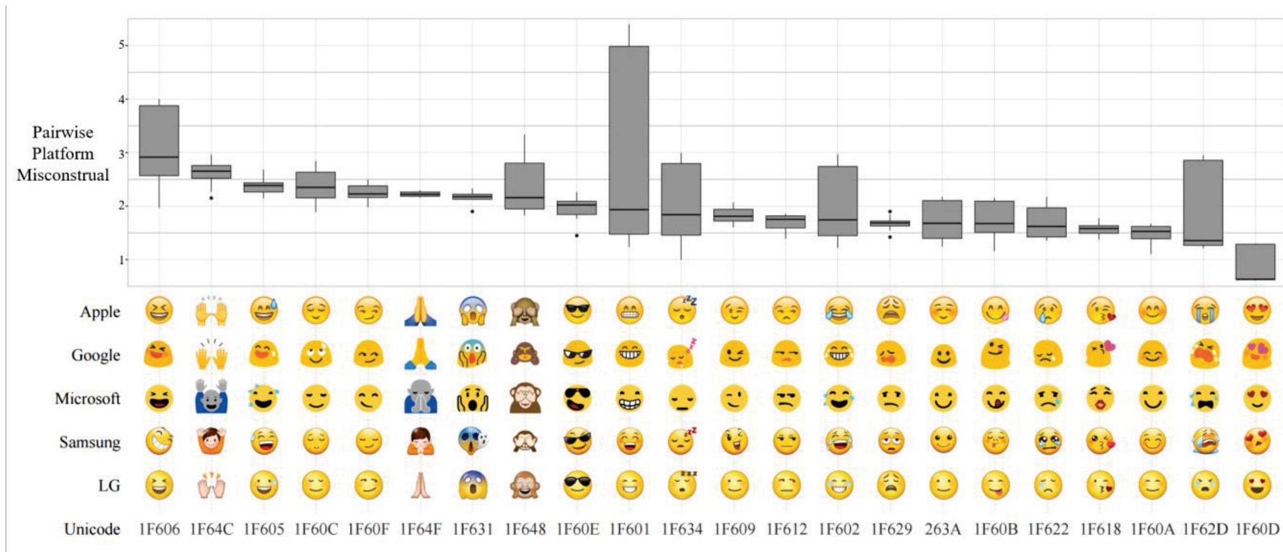

Figure 1. Across-platform sentiment misconstrual scores grouped by Unicode. Each boxplot shows the range of sentiment misconstrual scores across the five platforms. They are ordered by decreasing median platform-pair sentiment misconstrual, from left to right.

from the other platforms. For instance, the large range for "grinning face with smiling eyes" (U+1F601) reflects the very wide disagreement 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 we find "person raising both hands in celebration" (U+1F64C) in the top three most different renderings for four of our five platforms, suggesting some Unicode characters are simply more ambiguous than others, leading to within- and across-platform differences.

The results from RQ1 and RQ2 regarding interpretation of sentiment suggest that there are opportunities for misconstrual for both within-platform and across-platform renderings of emoji.

Semantic Analysis

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

Methods

We analyzed the free-text responses to Questions 1, 2, and 4 from our survey, which focused on the perceived meaning and use cases for the emoji. Here, we use a very similar technique to that presented above, adapted for text responses. For each participant's answer for each rendering, we 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, we ended up with 30 to 41 word vectors representing the responses of different participants. We 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." We compute overall difference in responses for a given emoji rendering as the average pairwise cosine distances of corresponding word vectors. This is similar to our within-platform sentiment misconstrual score above, so we will refer to this as our *within-platform semantic misconstrual* score. These values range from zero to one, increasing as participants use a greater variety of words in their responses, and 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, we present samples of aggregated responses in Table 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."

To answer our two research questions with regard to semantic interpretation, we ran a similar analysis as the one we did for sentiment. We first use the within-platform semantic misconstrual score described above to answer RQ1.

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

Table 2. 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.

We also computed *across-platform semantic misconstrual scores* of each Unicode character, mirroring the computation for our sentiment analysis. For each Unicode character (e.g., "face with tears of joy") and each pair of platforms (e.g., Apple and LG), we compute the pairwise word vector distances between the two sets of word vectors, and then take the average (e.g., an Apple-LG average word vector distance for the "face with tears of joy" emoji). We then computed the grand-mean (mean of these platform-pair average word-vector distances) to get the across-platform semantic misconstrual score for each Unicode character.

	Most/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)

Table 3. Top-3 and bottom-3 most differently described renderings. Higher values indicate greater response variation.

Results

RQ1 (Within Platform) for Semantics

Shown in Table 3, we 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 responses from the participants for a given rendering. Though the emoji 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.

RQ2 (Across Platform) for Semantics

Figure 2 shows the distribution of across-platform semantic misconstrual scores for all platform pairs (e.g., Google and Apple, Apple and Microsoft, etc.) for all emoji Unicode characters. We 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 platform-specific word vectors differed from one another, for each Unicode character. Indeed, we observe that there are statistically significant differences in the platform interpretations of Unicode 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 Unicode 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.

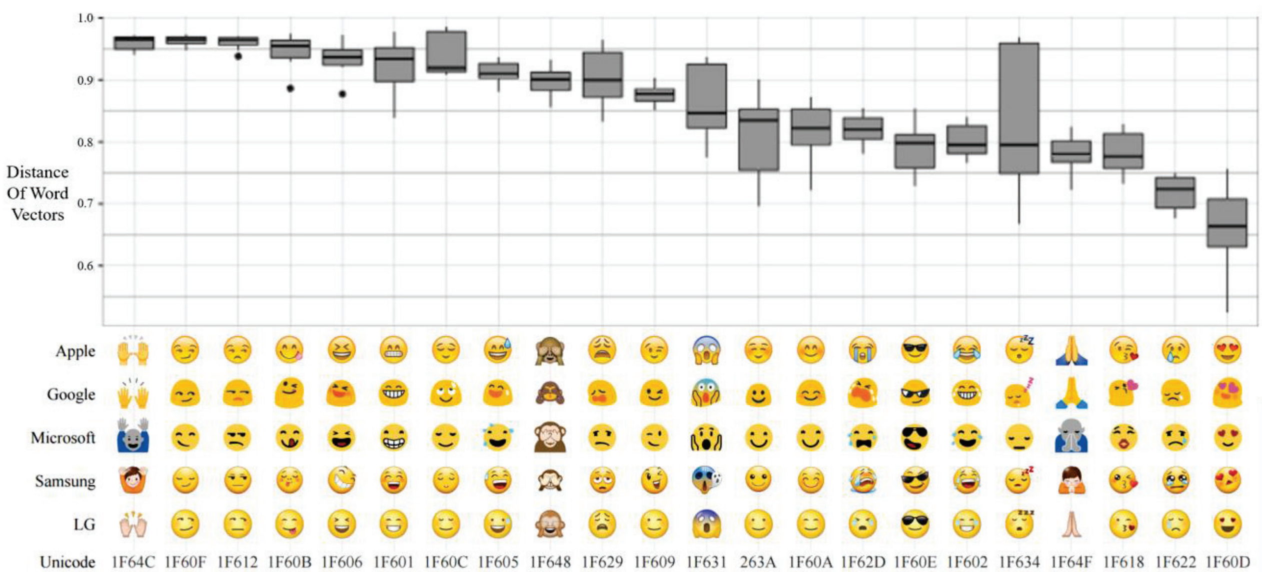

Figure 2. Across-platform semantic misconstrual scores grouped by Unicode. Each boxplot shows the range of semantic misconstrual scores across the five platforms. They are ordered by decreasing median platform-pair semantic misconstrual, from left to right.

Results Summary

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

RQ1: We find 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 opportunity for *within-platform misconstrual*. On our sentiment scale, only 4.5% of our renderings have an average misconstrual score below 1, and 40% have scores larger than 2, and our semantic analysis finds very few renderings are described the same way.

RQ2: We find that for both sentiment and semantic interpretations across platforms, there is disagreement. For a given emoji Unicode character (five renderings, one for each platform), there is clear opportunity for *across-platform misconstrual*. 9 of the 22 (41%) Unicode characters have sentiment distributions wider than one sentiment unit, and we see 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? We found that misconstrual was incrementally larger across-platform than within-platform. More specifically, the average across-platform sentiment and semantic misconstrual scores were 2.03 and 0.86, respectively (considering all across-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).

Discussion and Implications

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

Contextualizing Our Results in Theory

In the context of Clark's psycholinguistic theory of communication discussed above (Clark 1996), 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 interprets what he thinks it means, or develops his own construal. If Bill's interpretation differs from Abby's intended meaning, then Bill misconstrued Abby's communication. Our results suggest that people often interpret emoji in diverse fashions, potentially leading to situations like that of Abby and Bill. With discrepancy between the sender's and receiver's interpretations, the sender's intended meaning is not commonly understood by both of them, so the communication suffers. From our results, we 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 main 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, our results show that people's interpretations for a given emoji character vary more across multiple platform renderings than for a single platform's rendering. This implies that communication across platform is even more prone to misconstrual than within-platform.

At the end of the survey, we asked our 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.

Implications for Design

Our 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 we have shown, this does not mean that interpretation is standardized across platforms. Converging on emoji renderings across platforms rather than diverging (e.g. to maintain distinctive branding) may reduce the variation of interpretation and thus lower the likelihood of miscommunication.

However, in addition to across-platform challenges, we also observed that a great deal of the diversity in interpretations 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.

Future Work and Limitations

Though we studied 22 of the most popular anthropomorphic emoji, there are currently 1,282 total emoji Unicode characters (including non-anthropomorphic ones). Likewise, we studied 5 of the most popular mobile platforms, but there are at least 17 platforms with their own unique emoji renderings. We also only looked at one *version* of each platform's emoji even though people do not consistently use the same version of operating systems. 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 our study).

There are *many* different emoji renderings, and they all may be subject to differing interpretation. It would be infeasible to survey all of them and new ones are constantly emerging. Developing models to predict the sentiment and consistency 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.

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 often accompanied by surrounding text (e.g., in a text message). Researchers have found that emoticons can affect the interpretation of a message (Walther and D'Addario 2001; Lo 2008), but the parallel for emoji has not yet been explored. Other researchers have developed emoji sentiment classifiers based purely on the sentiment of text they appear in (Liu, Li, and Guo 2012; Novak et al. 2015), but this reflects interpretation solely of context and not the emoji themselves. It is an important direction of future work to explore people's interpretations of emoji with respect to the contexts in which they appear.

Another interesting avenue of future work lies in the potential for cultural differences in interpretation of emoji. Originating in Japan with global expansion, it is likely that emoji usage and interpretation is culturally dependent. Additionally, our approach to semantic analysis could be extended to use semantic relatedness measures, which would address challenges associated with vocabulary mismatch.

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. Psycholinguistic theory suggests that interpretation must be consistent between two people in order to avoid communication challenges. In this research, we 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 Unicode characters, we find disagreement in terms of both sentiment and semantics, and these disagreements only increase when considering renderings across platforms.

Acknowledgments

We thank the Mechanical Turk workers who participated in our survey. The National Science Foundation Graduate Research Fellowship Program under Grant No. 00039202 and a University of Minnesota College of Science and Engineering Graduate Fellowship supported this work. We would also like to thank our anonymous reviewers for their comments and suggestions for follow-up research.

References

Clark, H.H. 1996. *Using Language*. Cambridge University Press.

Davidov, Dmitry, Oren Tsur, and Ari Rappoport. 2010. "Enhanced Sentiment Learning Using Twitter Hashtags and Smileys." In *ACL '10*, 241–49. Association for Computational Linguistics.

Davis, Mark, and Peter Edberg. 2015. "Unicode Emoji." Technical Report 51.

Derks, Daantje, 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): 766–85.

Dimson, Thomas. 2015. "Emojineering Part 1: Machine Learning for Emoji Trends." *Instagram Engineering Blog*.

Huffaker, David A., and Sandra L. Calvert. 2006. "Gender, Identity, and Language Use in Teenage Blogs." *JCMC* 10 (2): 00–00.

Kelly, Ryan, 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*.

Liebman, Noah, and Darren Gergle. 2016. "It's (Not) Simply a Matter of Time: The Relationship Between CMC Cues and Interpersonal Affinity." In *CSCW '16*.

Liu, Kun-Lin, Wu-Jun Li, and Minyi Guo. 2012. "Emoticon Smoothed Language Models for Twitter Sentiment Analysis." In *AAAI '12*.

Lo, Shao-Kang. 2008. "The Nonverbal Communication Functions of Emoticons in Computer-Mediated Communication." *CyberPsychology & Behavior* 11 (5): 595–97.

Novak, Petra Kralj, Jasmina Smailović, Borut Sluban, and Igor Mozetič. 2015. "Sentiment of Emojis." *PloS One* 10 (12).

Park, Jaram, Vladimir Barash, Clay Fink, and Meeyoung Cha. 2013. "Emoticon Style: Interpreting Differences in Emoticons Across Cultures." In *ICWSM '13*.

Pavalanathan, Umashanthi, and Jacob Eisenstein. 2016. "Emoticons vs. Emojis on Twitter: A Causal Inference Approach." *AAAI 2016 Spring Symposium on Observational Studies through Social Media and Other Human-Generated Content*.

SwiftKey. 2015. "Most-Used Emoji Revealed: Americans Love Skulls, Brazilians Love Cats, the French Love Hearts." *SwiftKey Blog*. April 21.

Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics* 37, 2: 267–307.

Tossell, Chad C., 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): 659–63.

Walther, Joseph B., and Kyle P. D'Addario. 2001. "The Impacts of Emoticons on Message Interpretation in Computer-Mediated Communication." *SSCR* 19 (3): 324–47.