Benign or Toxic? Differences in Emoji Interpretation by Gender, Generation, and Emoji Type

Marina Zhukova[1] and Susan C. Herring[2]

[1]University of California, Santa Barbara
[2]Indiana University Bloomington

"*I must lead a sheltered life*" (Male Baby Boomer, when shown the message: "Meet me at the park at 6pm, fool 💣")

Abstract

Emoji usage, traditionally associated with playfulness and friendliness, now extends into toxic online behavior, encompassing threatening emojis, such as the bomb 💣, as well as positive emojis, such as a smile 😊, used to intensify toxic content. This study investigates the extent to which different demographic groups perceive emoji use differently, including potentially toxic emoji use. Through an online survey, participants rated combinations of text and emoji on scales of friendliness, playfulness, sarcasm, passive-aggression, offensiveness, and threat. The findings reveal differences in emoji perception by gender, generation, and emoji type. Notably, non-binary individuals and Gen Z respondents perceived the emoji-containing messages as more sarcastic, passive-aggressive, and threatening. All groups reacted most negatively to the bomb emoji and perceived the smiley face as friendly, but the smile and the thumbs up emojis were rated as more passive-aggressive by younger generations, while the fire and bomb emojis were rated more negatively by older generations. These findings have implications for the detection and moderation of toxic emoji use in online environments, as well as in legal contexts where emoji usage plays a role.

Introduction

Emojis have long been associated with positivity and friendly intentions. Herring and Ge-Stadnyk (2024) write that "emoji use generally connotes cuteness, playfulness, [humor,] and so forth at a meta-pragmatic level." In addition to positive uses, however, emojis are increasingly used in online spaces in problematic ways. These include cyberbullying (Matulewska & Gwiazdowicz, 2020), harassment (Goldman, 2022), and racist discourse (Matamoros Fernandez, 2018), as well as the use of emoji to mock, threaten, or give offense. We refer to these phenomena collectively as toxic emoji use. Emoji can potentially be toxic in two main ways. Emoji such as the water pistol 🔫 and bomb 💣 can themselves express violence or threat, while semantically neutral or positive emoji, such as the thumbs up 👍 and smiling face with smiling eyes 😊, can be used together with toxic text to emphasize its toxicity (Wagner & Yu, 2021).

As the level of toxicity of discourse in many online environments continues to rise (e.g., Casula et al., 2021; Kim et al., 2022), it is important to detect and moderate toxic emoji use along with other forms of abusive communication. Identifying potentially toxic emoji poses challenges, however. The meanings represented by emojis are often ambiguous (Częstochowska et al., 2022), as are the intentions underlying their use (Völkel et al., 2019). Moreover, emoji understanding has been shown to vary across user demographics, including among people of different genders and ages (Cui, 2022; Herring & Dainas, 2018, 2020). Meanwhile, younger generations of users are shifting the interpretation of some emoji, such that emoji that previously had positive connotations, such as the tears of joy and thumbs up emoji, are now interpreted negatively by some younger users as sarcastic or passive-aggressive (Katz, 2022). This leads us to ask: Do some demographic groups perceive certain emoji uses as toxic, while for other groups they are benign or even positive? If so, what are the implications of these differences for detection and moderation of toxic emoji use in online environments?

To address these questions, we conducted an online survey in which respondents of different genders and ages were asked to rate a neutral message and a potentially threatening message plus one of four emojis on scales of friendliness, playfulness, sarcasm, passive-aggression, offensiveness, and threateningness. Statistical and content analyses of the survey responses reveal two main findings. First, message content, message valence, context, and familiarity with the sender all significantly impacted how the emojis were interpreted, demonstrating that emoji understanding depends heavily on the surrounding text and the receiver's relationship with the sender. Second, consistent with research on non-toxic emoji use (e.g., Herring & Dainas, 2018, 2020), gender and age differences were found in the interpretation of potentially toxic emoji uses. Non-binary individuals and Gen Z respondents perceived the emoji-containing messages as more toxic overall, as indicated by their higher ratings for sarcasm, passive-aggression, and threat. All groups reacted most negatively to the bomb emoji and rated the smiley face as friendly. However, younger generations rated the smile and the thumbs up emojis as more passive-aggressive, while older generations rated the fire and bomb emojis more negatively, thus providing evidence for generational shift in emoji interpretation. Alongside these differences, our results reveal an overall positivity bias associated with emoji use.

Background

Toxic Emoji: Overview

Emojis are increasingly used in a variety of toxic ways. Like other forms of abusive language, they can be sexist and racist. One of the earliest published reports of problematic emoji use (Vocativ, 2015) called attention to the use of the eggplant emoji, which is often used to represent male genitalia, to sexually harass women. An example of racist emoji use is the appropriation of Facebook Reaction emoji and stickers to amplify Islamophobia by a far-right political party in Belgium, as reported by Matamoros Fernandez (2018).

Emoji are also used to express and amplify threats and aggression. Matulewska and Gwiazdowicz (2020) identified three uses of emoji to cyberbully hunters in Poland: 1) Negative emojis such as pouting face[1] 😡, angry face 😠, and nauseated face 🤢 were used to express hostility toward the hunters; 2) positive emojis such as love 💗, laughter 😆, and satisfaction 😊 were used to express support for people who cyberbullied others; and 3) positive emojis were used to ridicule, to give extra emphasis to messages formulated by cyberbullies, or to express happiness that something awful might happen to a cyberbullied person. Wagner and Wu (2021) classify emoji of the first type, which they call "power amplifiers in cyberbullying" (p. 950), into two broad categories: *human* emojis (including faces expressing strong negative emotion, such as the vomiting emoji, and hands making aggressive or obscene gestures) and *instrumental* emojis (including lethal weapons, sexist emojis, discriminatory emojis, and death threat emojis). These observations underscore that emojis can be toxic due to their negative semantics (e.g., 💣), as well as through positive amplification of negative message content (e.g., 'He will rip your cowardly heart out 😊' – implied: and it will give me pleasure).

Adolescents are frequent emoji users, and they are also especially likely to use emojis aggressively (Danesi, 2021). A growing number of legal cases revolve around the use of emoji, many of them involving juveniles. Danesi (2021) cites the case of a California youth who was convicted in 2015 of making a criminal threat to shoot up his school in a tweet stream that included laughing and clapping emojis. Another student in Nebraska was disciplined for using the fire emoji as an apparent threat of a violent attack on his school. And in Virginia, a 12-year-old girl was charged with threatening her school through an Instagram post that said: "meet me in the library Tuesday 🔫 🔪 💣" (Ferreira, 2016).

Violent emojis such as the knife, gun, skull (symbolizing death), and bomb are also used by gang members, including to intimidate witnesses from testifying against them in court (Danesi, 2021). Offenders often use the defense that they were just kidding, relying on the prevalent view of emojis as playful and harmless. However, the contexts in which such emojis are shared indicate that the senders' intentions are often far from harmless. Of the 166 US court opinions that referenced emojis and emoticons in 2021, 21 involved sexual predation, 19 involved employment discrimination, and nine involved murder (Goldman, 2022). Danesi (2021) concludes that "the darker side of human interaction is manifesting itself more and more via emoji textuality."

Detection of Emoji Toxicity

Efforts to automatically detect and filter out toxic content in public online spaces are increasingly focusing on emojis. Wiegand and Ruppenhofer (2021) found that including negatively valenced emojis such as the middle finger, pile of poo, vomiting emoji, skull, and water gun in training models helped disambiguate potentially abusive lexical items, such as profane terms, in English, German, and Portuguese Twitter tweets. Kim et al. (2022) improved detection of toxic messages in Twitch chat by identifying emotes (Twitch-specific emoji) used to replace letters in racist and other abusive or offensive words, a practice used to evade algorithmic moderation. Using a human-

and-model-in-the-loop approach, Kirk et al. (2021) created the HatemojiBuild dataset consisting of 5,912 adversarial examples with which to train models to detect emoji-based hate.

Humans are better than machines at identifying when emoji use is intended as abusive, as opposed to ironic or playful (Kim et al., 2022). Automatic methods are needed, though, because toxic messages are too numerous for human moderators to manage (Kim et al., 2022). Yet automatic detection of toxic emoji poses challenges. Absent context, many, if not most, emoji are semantically ambiguous (Częstochowska et al., 2022). The pragmatic intention behind emoji use can also be ambiguous (Völkel et al., 2019). Emojis can be used in place of threatening words to avoid detection as abusive language (Ferreira, 2016), to soften potentially threatening meanings of a message (McMahon & Kirley, 2019), to indicate irony or sarcasm (Subramanian et al., 2019), or simply to lighten the tone of a message (Danesi, 2016). Moreover, the context in which an emoji is used – on what platform, on what topic, from what kind of a sender, and so forth – determines to a significant extent whether messages are meant to be taken literally, ironically, sarcastically, or otherwise (Cui, 2022; Wagner et al., 2021). Such context is often unavailable or not included in algorithmic approaches.

Differences in Emoji Perception

An additional challenge for toxic emoji detection – and for abusive language detection more generally – is that "toxic" messages may be interpreted differently by different receivers. What is perceived as offensive or threatening by some individuals or demographic groups may be benign or even amusing to others. This is a largely unexplored phenomenon as regards toxic emoji use. However, research has shown that the interpretation of (benign) emoji can vary according to the gender and age of the receiver. Herring and Dainas (2018, 2020) conducted an online survey to determine how emojis in comments modified from public Facebook groups were interpreted by different genders and ages. They found that individuals who self-identified as "other" gender interpreted emojis significantly differently from self-identified females and males. No overall differences were found between the interpretations of females and males, even though younger females, especially, like and use emojis more than males do, and use them in different ways (e.g., Chen et al., 2018; Prada et al., 2018). Similarly, Jones et al. (2020) found no female-male gender difference in valence ratings of positive face emojis such as 😂 and 😍. However, women rated negatively valenced faces such as 😈 and 🤬 as significantly more negative than men did, evidence of an "emotional negativity bias" for women in facial emotion processing (Jones et al., 2020). Herring and Dainas (2020) also found age differences in emoji understanding: Older respondents to their survey were more likely to interpret the emojis literally compared to younger respondents, who recognized more nuanced, non-literal meanings of emojis. These generational differences were especially pronounced between younger females and older males.

Certain emojis are used and interpreted differently by younger generations. An example that has been the subject of several scholarly studies is the Chinese smile emoji with round eyes 🙂. An et al. (2018) found that 16-to-25-year-old Chinese users interpreted the smiling emoji as negative or neutral in valence, while users over 36 years old associated it with positive sentiment. In another

study involving Chinese speakers that controlled for the age of the sender as well as the age of the receiver, Cui (2022) found that the smile emoji was more likely to be interpreted as sarcastic or even mean by younger participants when sent by young senders, but older participants tended to perceive the smiling emoji as positive regardless of sender age. These differences align with findings that young adults are more likely to use, and are better at identifying, sarcasm compared to older adults (Garcia et al., 2022; Rothermich et al., 2021).

Sarcastic uses and interpretations have also been reported anecdotally for Western emoji that were previously considered positively valenced – and that are still positive for older generations of users. For example, Katz (2022) and Steinberg (2022) report that the thumbs up emoji is perceived as sarcastic or passive-aggressive and avoided by many younger users. The tears of joy 😂 is also avoided by Gen Z, or they may use it ironically to represent older people's usage (Yurieff, 2021). The generic smiley face 🙂 has been variously reinterpreted by Gen Z as fake, ironic (Dictionary.com, 2022), passive-aggressive, or cold (Abril, 2022). As for the red heart, its use in the expression "No ❤️" is an example of a "fairy comment" in which soft emoji such as sparkles, butterflies, and hearts are used to convey extreme sarcasm, for example, by teens dissing politicians (Sanjay, 2020). A Gen Z-er quoted by Capobianco (2022) reported that her generation uses the smiling face with hearts 🥰, which for older generations is a loving image, as "a passive-aggressive way of bullying someone online."

These observations raise the very real possibility that different genders and generations perceive emoji-based toxicity differently, including offensive and threatening emoji use. We might expect, for example, Gen Z to be more likely than older generations to interpret emoji such as smiles and the thumbs up as ironic or passive-aggressive. But would Gen Z interpret violent emoji such as bombs, fire, and guns ironically, and would they consider such uses non-threatening? As regards gender, we might expect women more than men to interpret toxic emoji use as negative, consistent with the female negativity bias in facial emotion processing identified by Jones et al. (2020). Relatedly, females and nonbinary individuals might be more sensitive than males to potentially toxic emoji usage, given that women and LGBTQ+ people are often targets of online harassment (Jane, 2017; Scheuerman et al., 2018). Thus, in this study we ask: To what extent do different demographic groups (dis)agree on what constitutes toxic emoji use, and how does this vary according to the particular emoji used?

Methods

Data Collection

Materials

We designed an online survey using Qualtrics to investigate how different genders and generations interpret messages containing emoji. We chose a survey methodology to gain a broad understanding of perceptions of emoji usage across the demographic groups, as surveys allow

larger and more representative samples to be recruited than is possible through methods such as experiments or interviews. Our main focus in this study is emoji use that is potentially threatening because of the message content, the emoji, or a combination of both.

As stimuli, we created 24 combinations of emoji, message content, and message valence. We chose four emoji that are in common use on social media. Two – the smiling face with smiling eyes 😊 and the bomb 💣 – are relatively unambiguously positive and negative, respectively. The other two – thumbs up 👍 and fire 🔥 – we expected to be more ambiguous, based on reports in the literature that those emoji are interpreted differently by different generations (Abril, 2022; Capobianco, 2022; Dictionary.com, 2022; Yurieff, 2021). To convey different degrees of potential threat, we constructed two sentences, a neutral one: "*I'll meet you at the park at 6 pm*" and a potentially threatening one: "*You're going to regret that.*" The first sentence was modeled after a message used by a 12-year-old girl to threaten her school that said: "meet me in the library Tuesday 🔫 🔪 💣" (Ferreira, 2016). Although neutral by itself, in combination with a threatening emoji, the message could be interpreted as threatening (and was, in the case of the schoolgirl). In contrast, the second sentence could constitute a threat by itself, due to the presence of a 2nd-person pronoun, future modality, and the suggestion of harm in the word 'regret,' which are common features of threatening messages (Bojsen-Møller, 2020).[12] We expected that adding a threatening emoji, in addition to making a threatening interpretation more likely, would amplify the threat in the message (Wagner & Wu, 2021), but that a positive emoji, paradoxically, could have a similar effect (Matulewska & Gwiazdowicz, 2020).

To introduce further nuance, we created three valenced forms of each sentence: positive, negative, and neutral, by adding an enthusiastic interjection ('great' or 'OMG') at the beginning to make the sentence more positive, by adding an insulting epithet ('fool') at the end to make it more negative, and by not adding anything to the sentence for the neutral version. This was done to examine the effects of emoji use on the meaning of the sentences with different sentiments, for example, whether adding the bomb emoji would cause a positively-valenced sentence to be interpreted as toxic. We also designed two fillers sentences in which a statement was randomly paired with an emoji: "*Well, you are entitled to your opinion* 🥺" and "*Well, you are entitled to your opinion* 👉👈," and one message for an attention check: "*Going to take the new puppy for a walk,*" followed by one of the following emoji: 😊, 💕, ❤️, 😍. The stimuli, fillers, and attention checks are listed in Appendix A.

Survey Design

To be eligible for participation, individuals had to 1) reside in the US; 2) consider themselves fluent in English; and 3) be at least 18 years old. At the beginning of the survey, participants were shown the study information sheet and the consent form. To qualify for the study, the participants needed to select either "I am a native speaker of English" or "I am able to use English in academic contexts/at work" in response to the question "How well can you speak and understand English?". Participants' consent was collected by verifying that they clicked on the button "I agree to respond

to this survey." The survey started with demographic questions about age (in years), gender (male, female, non-binary, prefer not to say, please specify), ethnicity, and education level.

The participants were then informed that they would see a series of potentially similar text messages and asked to read each message and rate it on the scales provided. They were instructed to evaluate the messages as they would personally interpret them if they saw them online or on their phone.

The messages were shown to the participants in the following order: 1) the first filler, 2) three randomly selected stimuli, 3) the second filler, 4) attention check, and 5) another three randomly selected stimuli. The distribution of stimuli was both equal and randomized to maintain a balanced and unbiased presentation of stimuli to the participants.

For each of the fillers and stimuli, participants were asked, "In what context is a message like this most likely to appear?" The choices provided were: in a private message or email from a friend; in a private message or email from a stranger; in a comment on a social media post by a friend; in a comment on a social media post by a stranger; in a comment on a discussion forum like Reddit; in a comment below a news article;[3] other. If participants thought the message would appear in some other context, they were asked to type it in the text box provided. After that, they were asked to rate each message as to what degree they perceived it to be playful, friendly, sarcastic or ironic, passive-aggressive, offensive, or threatening along 5-point Likert scales (not at all, a little, to some extent, moderately, very much, not sure). For the attention check, participants were presented with four different versions of the attention check sentence and asked to select the version that was followed by the Red Heart emoji.

The last part of the survey consisted of questions about how often the participants use emoji in the messages they post to social media and how often they use social media.

Procedures

The survey was distributed on social media (Twitter, Facebook, Instagram, LinkedIn) and through department mailing lists at the first author's institution using snowball sampling. The data were collected during February and March 2023. Out of 340 responses received, 273 responses met the eligibility criteria and were used for further data analysis.

Participants

We consolidated our participant sample into four self-identified gender categories: females, males, non-binary individuals, and individuals who prefer not to disclose their gender, and four age groups: Baby Boomers (born between 1946 and 1964), Gen X (1965-1980), Millennials (1981-1996), and Gen Z (1997-2005). Three "please specify" responses were incorporated into the closest relevant categories. The participants' gender and age distribution is summarized in Table 1.

Gender/ Generation	Female	Male	Non-Binary	Prefer not to say	Total
Gen Z	47	30	14	6	97
Millennials	54	19	18	3	94
Gen X	33	10	0	1	44
Baby Boomers	25	12	1	0	38
Total	159	71	33	10	273

Table 1. Gender and age distribution of survey participants

Although we accounted for English proficiency in our screening criteria, we did not differentiate between native and non-native speakers in our analysis, since the focus of our study was on perceptions across generational and gender groups. All participants met the minimum English proficiency threshold required to understand and respond to the survey items appropriately.

Data Analysis

Statistical Methods

Context and Familiarity

We used cumulative link mixed models (*clm* package in R Studio) to examine how various predictor variables influenced respondents' perceptions of the context of the message (public vs. private) and the familiarity of the sender (friend vs. stranger). We tested the effect of the message content (MESSAGE: park (reference level) vs. regret; deviation coding), the type of emoji (EMOJI: bomb (reference level) vs. smile vs. fire vs. thumbs up, deviation coding), gender of the participant (GENDER: female (reference level) vs. male vs non-binary vs. prefer not to say; deviation coding), generation of the participant (GENERATION: Gen Z, Millennials, Gen X, Baby Boomers; Reverse Helmert coding), and message valence (VALENCE: neutral (reference level), positive, negative; deviation coding) on the perceived message context (CONTEXT; public vs. private) and perceived sender of the message (FAMILIARITY; friend vs. stranger). The Reverse Helmert coding scheme compares each level to the mean of the previous levels, allowing us to compare each generation to the mean of the previous generations, rather than the reference level, and thereby better accounting for differences across generations. The item-level variability was included as a fixed effect. A random intercept per participant was added.

The first model (CONTEXT ~ MESSAGE + EMOJI + GENDER + GENERATION + VALENCE + (1|PARTICIPANT)) predicted perceived message context (public or private) as a function of message content (park or regret), emoji type (💣, 😊, 👍, 🔥), gender (female, male, non-binary, prefer not to say), generation (Gen Z, Millennials, Gen X, Baby Boomers), and message valence (positive, negative, neutral). The second model (FAMILIARITY ~ MESSAGE + EMOJI +

GENDER + GENERATION + VALENCE + (1|PARTICIPANT)) predicted perceived familiarity of the message sender (friend or stranger) using the same set of predictor variables.

In assessing the statistical significance of our findings, we considered a threshold of $p < 0.05$ as indicating significance. Model coefficients can be found in Appendix B (Tables A and B).

Friendliness, Playfulness, Sarcasm, Passive-Aggressiveness, Offensiveness, and Threat

Cumulative link mixed models (*clm* package in R Studio) were also used to examine the relationships between predictor variables MESSAGE, CONTEXT, FAMILIARITY, VALENCE, GENDER, GENERATION, EMOJI and participants' perception of friendliness, playfulness, sarcasm, passive-aggressiveness, offensiveness, and threat. These relationships were tested using the same values for the predictor variables as described in the first paragraph of the *Context and Familiarity* section.

For the interaction GENDER x GENERATION, we looked at the subset of the data for the youngest two generations, Gen Z and Millennials, as we had enough participants in all the gender categories only for these generations (GENERATION: Gen Z (reference level) vs. Millennials, deviation coding). The dependent variables (FRIENDLY, PLAYFUL, SARCASTIC, PASSIVE-AGGRESSIVE, OFFENSIVE, THREATENING) were measured on 5-point Likert scales (not at all, a little, to some extent, moderately, very much). "Not sure" was also an option, but "not sure" responses were not included in the statistical analysis. The item-level variability was included as a fixed effect. A random intercept per participant was added.

The models were fitted, and model significance was evaluated based on p-values, considering a threshold of $p < 0.05$ as indicating significance. To assess the effects of GENDER, GENERATION, EMOJI, and their interactions on the perception of the messages with emoji, we constructed separate models for each of the dependent variables, with the following predictors:

1. CONTEXT + FAMILIARITY + MESSAGE + VALENCE + GENDER + GENERATION + EMOJI + (1|PARTICIPANT)
2. CONTEXT + FAMILIARITY + MESSAGE + VALENCE + GENDER x GENERATION + (1|PARTICIPANT)
3. CONTEXT + FAMILIARITY + MESSAGE + VALENCE + GENDER x EMOJI + (1|PARTICIPANT)
4. CONTEXT + FAMILIARITY + MESSAGE + VALENCE + GENERATION x EMOJI + (1|PARTICIPANT)

Content Analysis

The survey included an open-ended response option for each message regarding the context in which the message was most likely to appear. We collected these comments and conducted content analysis by manually classifying each comment according to a set of themes that emerged from

the data. The classification and coding were done by both authors together. We calculated the frequency of the themes and their distribution according to the respondents' gender and generation. However, since not many survey respondents provided open-ended responses, the theme frequencies when broken down by respondent demographics were too small to report as quantitative findings. Instead, we report the overall frequencies of the themes and illustrate them with quotes from the open-ended responses, noting when certain themes were mentioned more often by a particular demographic group.

Results

In this section, we first present the statistical results for the contextual and message content variables, followed by the statistical results for gender, generation, and emoji type. The content analysis results are then presented and illustrated.

Statistical Analysis

Context and Familiarity

Survey respondents were asked to indicate in what kind of context they were most likely to encounter each message. The type of emoji and the content and valence of the message all impacted whether the message was perceived as having been sent in a public or private context. The regret messages were more often seen as public. The fire emoji was also more often perceived as public, while the smiley face and the thumbs up emoji made messages much more likely to be perceived as private. Negatively valenced messages were associated with increased perception that they were sent publicly; however, positive valence did not have a significant effect. Nor did the gender or generation of the respondents significantly influence perceptions of the message context as public or private. See Table A1 in Appendix B.

When a message was perceived to be public as opposed to private, there was a significant decrease in the perception of the message's friendliness ($p < .001$) and a significant increase in the perception that it was sarcastic or ironic ($p < .01$), passive-aggressive ($p < .001$), and threatening ($p < .05$). See Figure A1 in Appendix B. The perceived publicness of the message did not significantly influence perceptions of playfulness or offensiveness.

The results for the familiarity of the sender are similar to those for publicness of the context. Regret messages were more often perceived as coming from a stranger. The presence of the fire emoji was also more often perceived as coming from a stranger, whereas messages with the smiley face were much more likely to be perceived as coming from a friend. Negatively valenced messages were perceived as having been sent by a stranger, although positive valence did not have a significant effect. The gender and the generation of respondents did not significantly influence perceptions of messages as originating from a friend or a stranger. See Table A2 in Appendix B.

Messages perceived as coming from strangers as opposed to friends were perceived as significantly less friendly, playful, and sarcastic or ironic (p < .001), and significantly more passive-aggressive, offensive, and threatening (p < .001). See Figure A2 in Appendix B.

Message Content: Neutral vs. Threatening

The content of the message also significantly influenced how it was perceived. Overall, compared to the neutral message "*I'll meet you at the park at 6 pm,*" the potentially threatening message "*You're going to regret that*" was less often associated with a private context (p < .001) and less often perceived as coming from a friend (p < .001).

Also, when the message content was "*You're going to regret that,*" respondents rated the message significantly lower on playfulness and friendliness (p < .001) and significantly higher on sarcasm or irony, passive-aggression, offensiveness, and threat (p < .001). The Likert item results broken down by message content are visualized in Figure A3 in Appendix B.

Message Valence: Positive, Negative, Neutral

Messages with negative valence were more often perceived as being sent publicly and from a stranger (p < .05). Compared to neutral messages, messages with positive valence were more often rated as friendly and playful (p < .001) and less often rated as threatening (p < .05). Conversely, messages with negative valence were more often rated as offensive (p < .001) and threatening (p < 0.001).[4] Message valence did not play a significant role in the perception of passive-aggression. The significant results are shown in Figure A4 in Appendix B.

Gender

Compared to females, non-binary individuals rated the messages overall as significantly more sarcastic or ironic (p < .01), as did individuals who preferred to not disclose their gender (p < .05). In addition, non-binary participants rated messages as marginally more passive-aggressive and threatening, compared to females (p < .09), although the effect was not significant. Respondent gender did not influence perceptions of friendliness, playfulness, or offensiveness. See Figure 1. (Note: Only significant results are presented in the figures below and in Appendix B. The order of the variables in the figures reflects the coding of the variables for the statistical analysis; the reference level variable is always displayed on the left.)
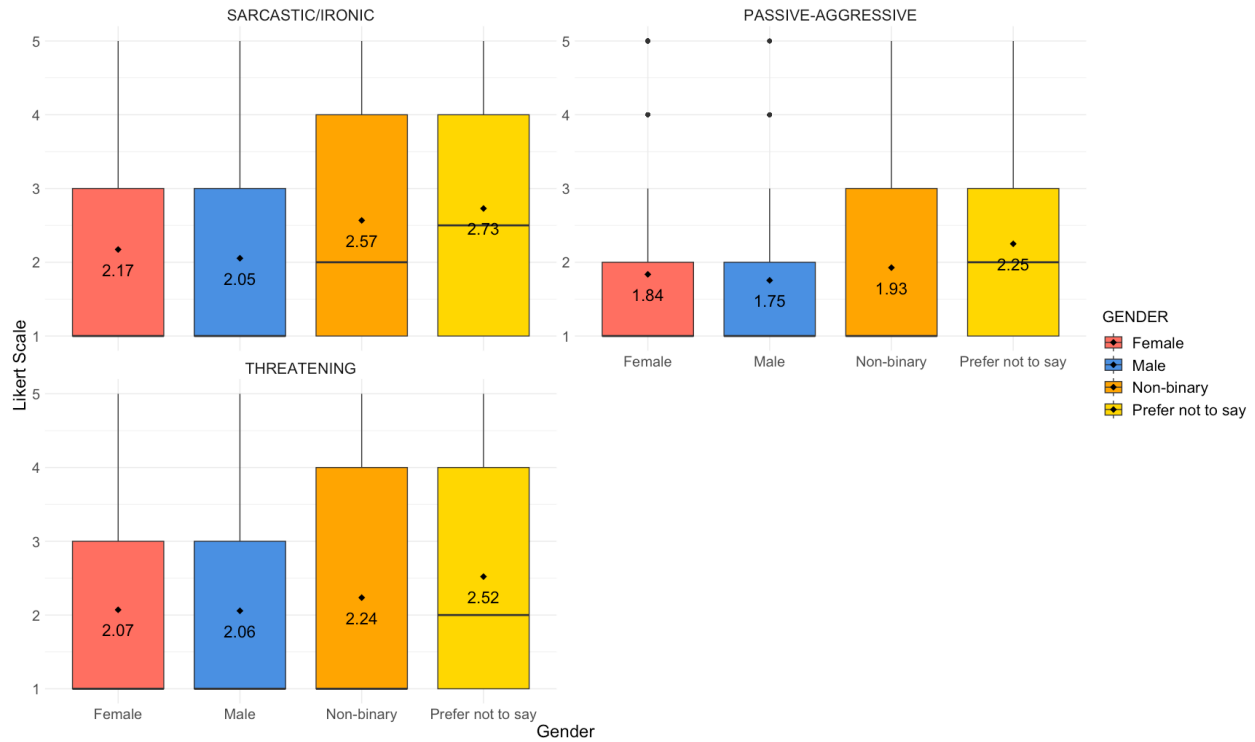
Figure 1. Likert Items by Participant Gender: Female vs. Male vs. Non-binary vs. Prefer not to say. The median is represented by the horizontal line, and the mean is represented by the black diamond in each box.

Generation

Gen Z respondents perceived messages as significantly more sarcastic or ironic overall ($p < .05$), compared to older generations. Millennials interpreted messages as more offensive ($p < .05$) than Gen Z and less offensive that Gen X. Compared to the younger respondents, Gen X respondents interpreted messages as less passive-aggressive ($p < .05$). Baby Boomers interpreted the messages as more offensive compared to younger generations ($p < .001$). These patterns are visualized in Figure 2.
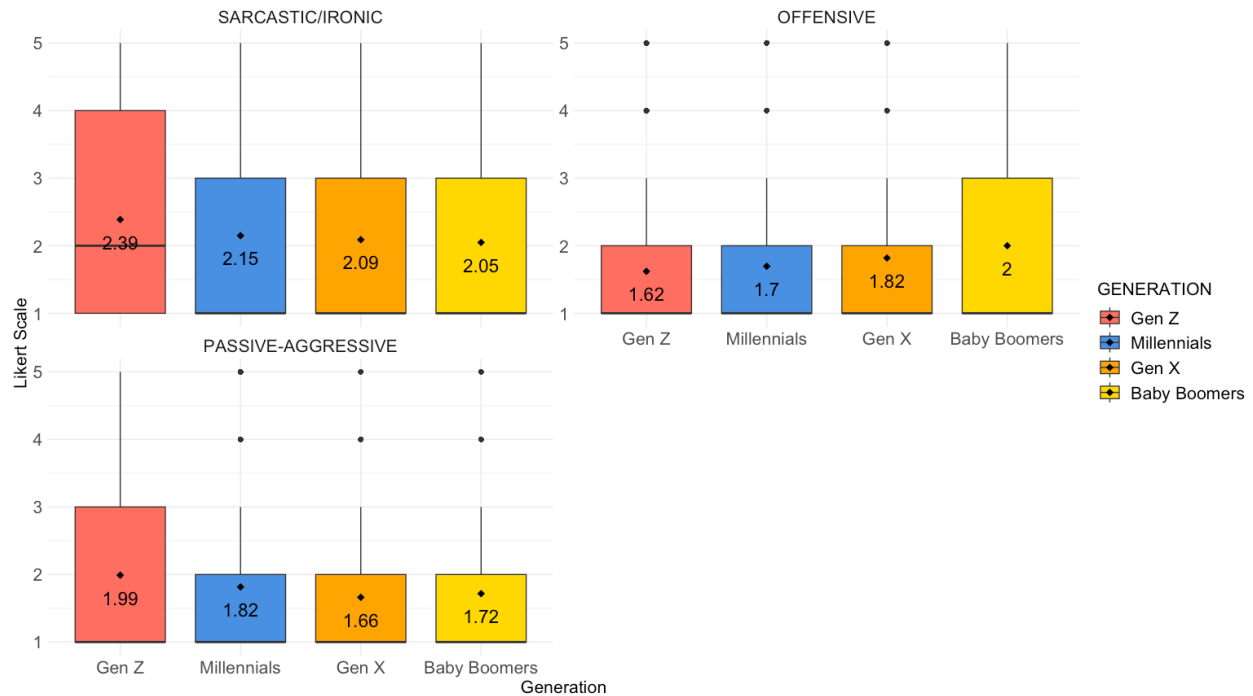
Figure 2. Likert Items by Participant Generation: Gen Z vs. Millennials vs. Gen X vs. Baby Boomers. The median is represented by the horizontal line, and the mean is represented by the black diamond in each box.

Gender x Generation

No significant differences were found for Gender x Generation effects between the Gen Z and Millennial generations, for which we had a sufficient number of participants to conduct the analysis. Marginally significant gender differences were found within the Millennial generation, however. Millennial males rated messages as less offensive than Millennial females did overall ($p < .06$).

Emoji Type

Emoji type was associated with significant differences in message interpretation. Among the four emojis, messages with the smile emoji were more likely to be rated as friendly ($p < .001$) and less likely to be rated as offensive or threatening ($p < .001$) compared to the bomb emoji. The bomb emoji elicited more negative reactions: Messages containing it were rated as significantly less friendly and significantly more offensive and threatening. Messages with the fire emoji were also rated as more friendly and playful and less offensive and threatening than the bomb emoji ($p < .001$). Moreover, the presence of the fire emoji significantly decreased ratings of sarcasm and irony ($p < .01$), while the smile emoji was ranked highest of the four emojis for sarcasm and irony ($p < .05$). Messages with the thumbs up emoji were perceived as more passive-aggressive ($p < .05$) compared with the bomb emoji. These findings are illustrated in Figure 3.
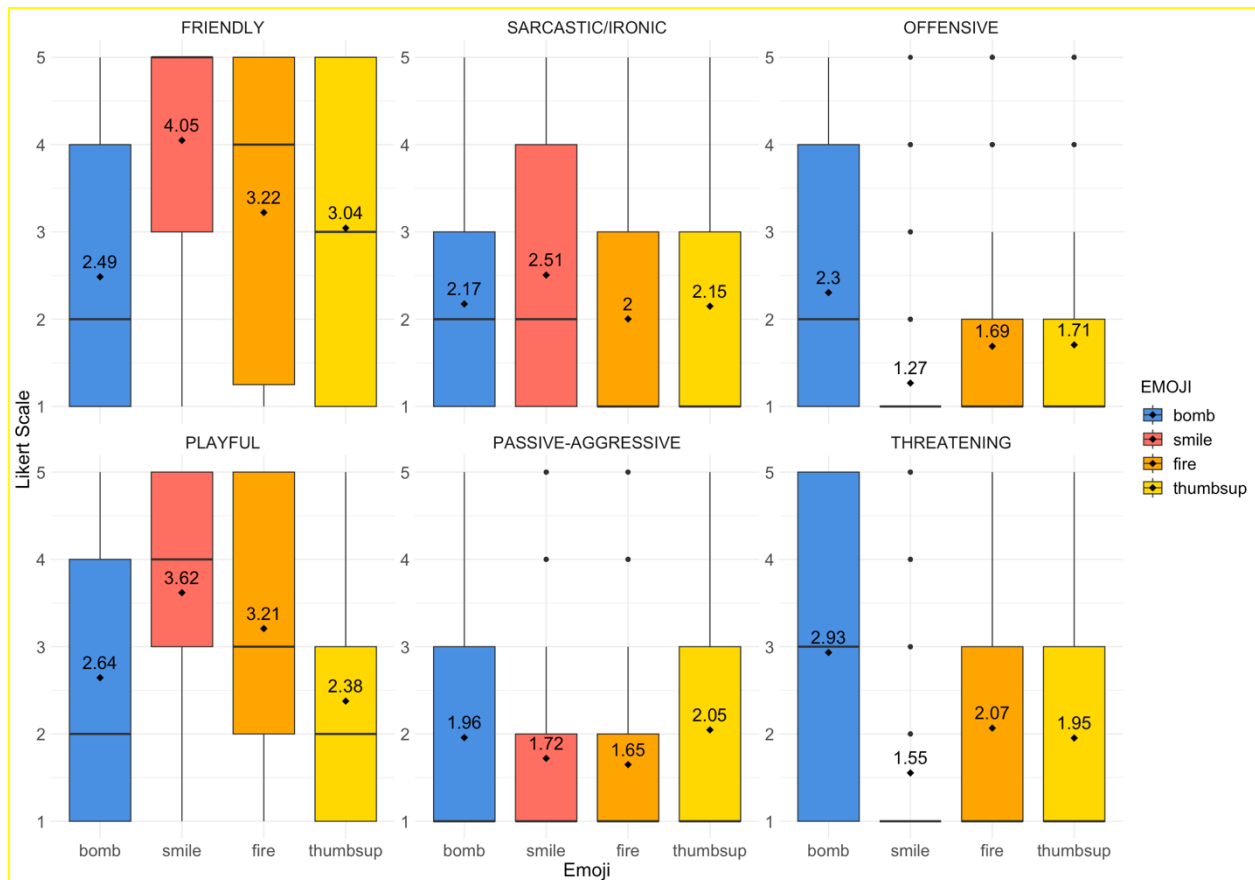
Figure 3. Likert Items by Emoji Type: Bomb vs. Smile vs. Fire vs. Thumbs Up. The median is represented by the horizontal line, and the mean is represented by the black diamond in each box.

Emoji x Gender, Emoji x Generation

Only the thumbs up emoji showed significant associations with participant gender. Males were more likely than females to perceive a message with the thumbs up emoji as sarcastic ($p < .01$), and individuals who preferred not to disclose their gender were less likely than females to perceive a message with the thumbs up emoji as offensive ($p < .01$).

In contrast, numerous significant associations were found between emoji type and participant generation. Compared to the older generations, Gen Z rated messages with the fire emoji as more friendly ($p < .05$), and they rated messages with the thumbs up as less friendly ($p < .05$). The fire emoji was rated as less friendly by older generations ($p < .05$). These results show a linear trend: The older the generation, the less friendly the fire emoji – and the more friendly the thumbs up – was perceived to be. Figure 4 shows the ratings for "friendly" by generation and emoji type.
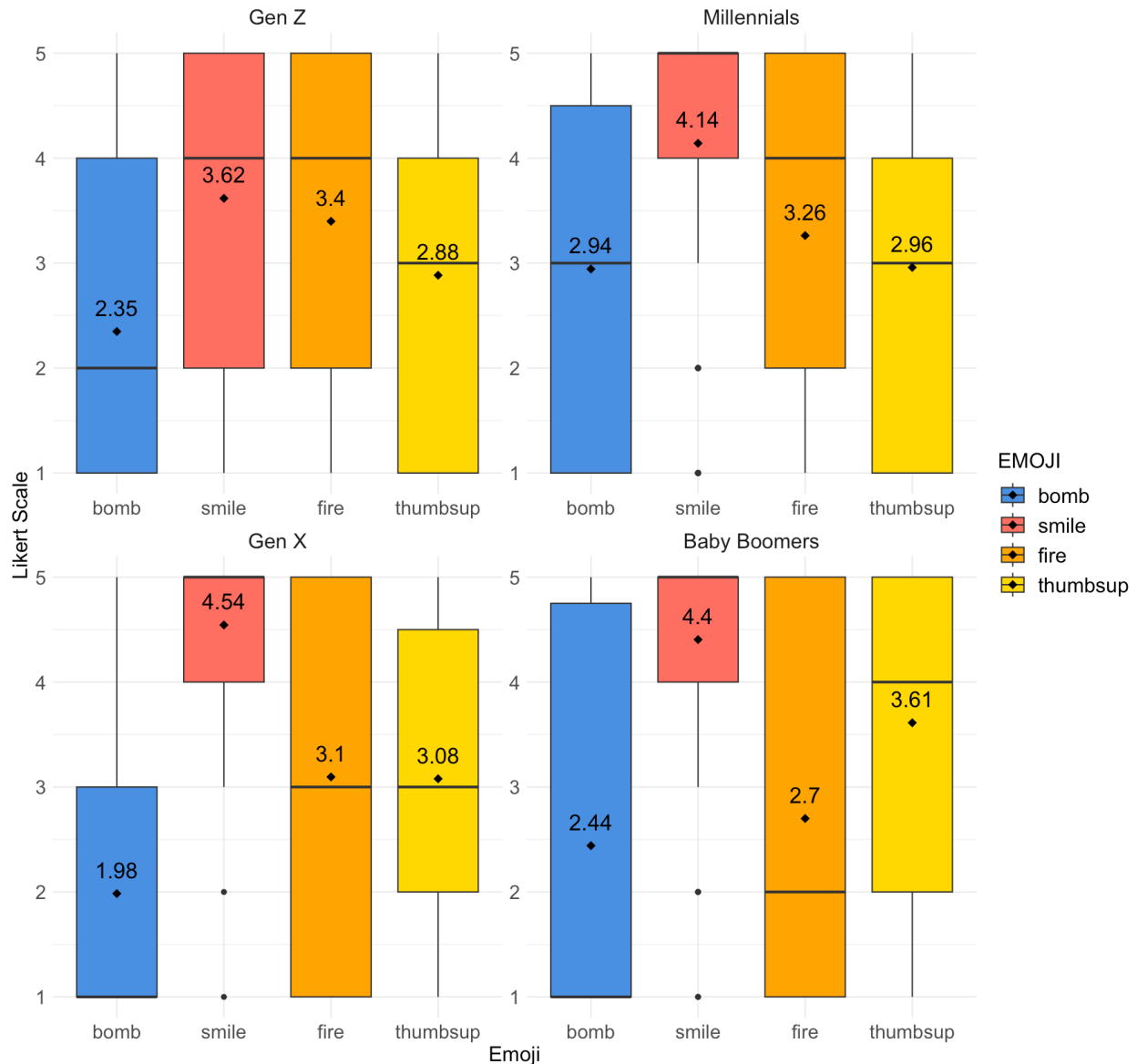
Figure 4. Likert Item for "Friendly" by Generation and Emoji Type. The median is represented by the horizontal line, and the mean is represented by the black diamond in each box.

As regards "playful" ratings by generation and emoji type, messages with the smile emoji ($p < .05$) were perceived by Gen Z and Millennials as significantly less playful compared to the older generations. In contrast, messages with the fire emoji were perceived by Gen Z and Millennials as significantly more playful compared to Gen X ($p < .05$). Baby Boomers rated messages with the thumbs up emoji as more playful than any of the younger generations did ($p < .01$). Excluding millennials, the playfulness results for the fire emoji and the thumbs up emoji exhibit linear trends in opposite directions, similar to the patterns noted above for friendliness. See Figure 5.
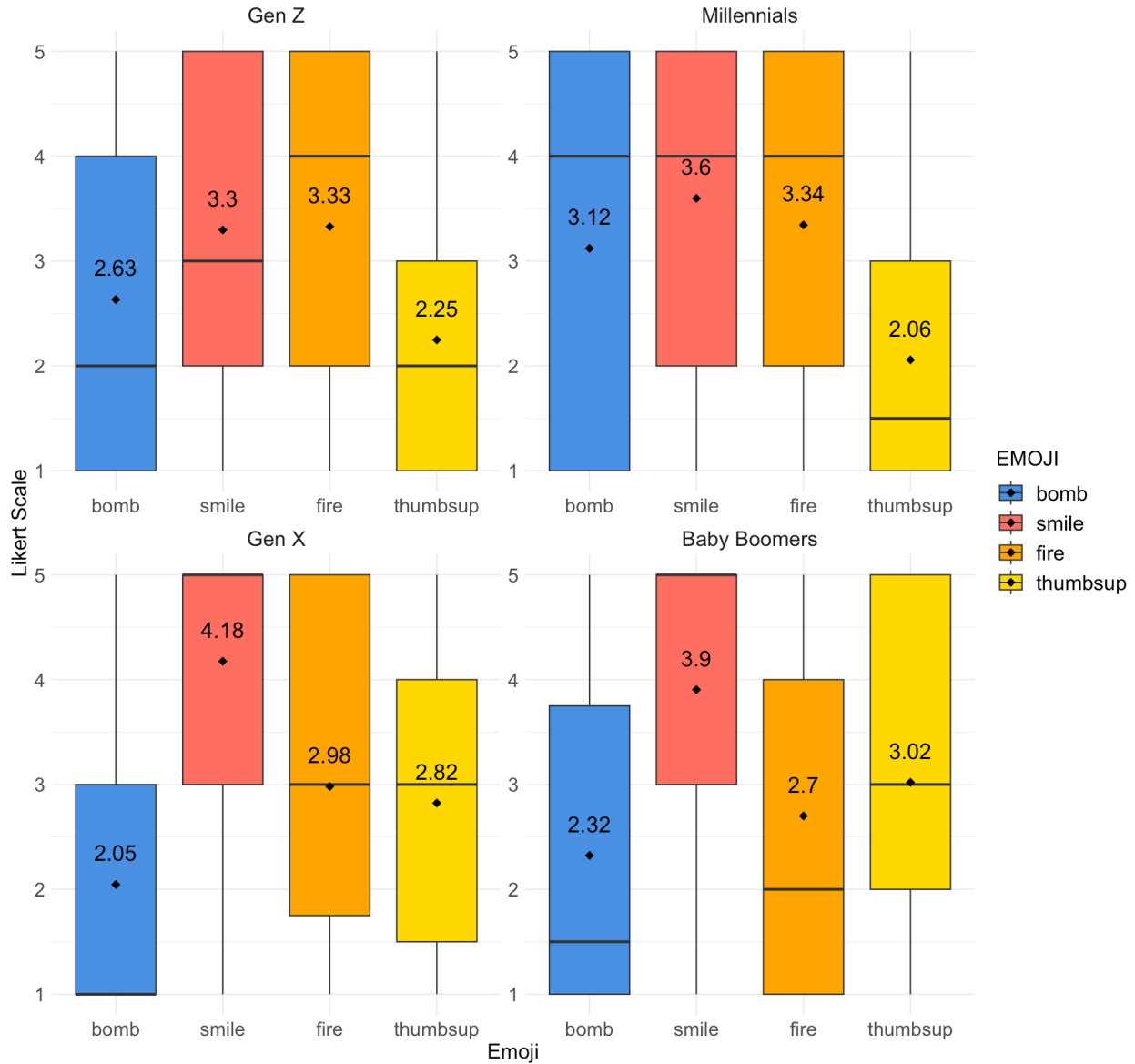
Figure 5. Likert Item for "Playful" by Generation and Emoji Type. The median is represented by the horizontal line, and the mean is represented by the black diamond in each box.

"Sarcastic or ironic" ratings by generation and emoji type showed less significant variation. Notably, however, Baby Boomers rated messages with the thumbs up emoji as less sarcastic or ironic compared to the younger generations ($p < .05$). See Figure 6.
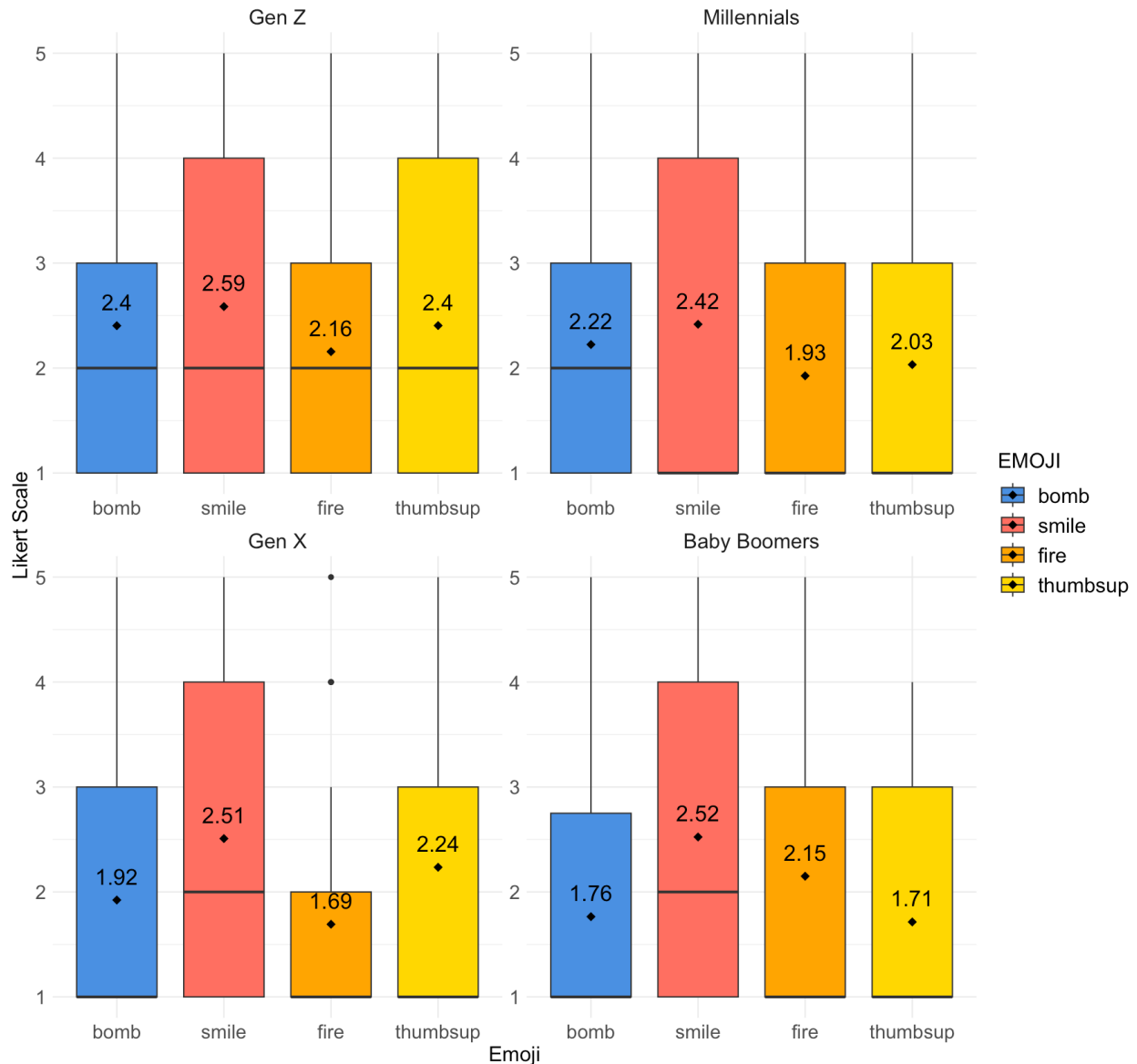
Figure 6. Likert Item for "Sarcastic or Ironic" by Generation and Emoji Type. The median is represented by the horizontal line, and the mean is represented by the black diamond in each box.

"Passive-aggressive" ratings showed clear generational differences. Baby Boomers rated messages with the bomb emoji as more passive-aggressive (p < .05) than did the younger generations, while rating messages with the smile emoji as significantly less passive-aggressive (p < .001). Conversely, Gen Z respondents evaluated messages with the smile emoji as significantly more passive-aggressive compared to Millennials (p < .05), while messages with the bomb emoji were evaluated as less passive-aggressive compared to the three older generations. Gen Z also rated messages with the fire emoji as more passive-aggressive than Millennials did (p < .05), while Millennials rated the same messages as more passive-aggressive than Gen X did (p < .01). Marginally significant differences (p < .06) were associated with the perception of the thumbs up emoji across generations: Gen Z and Millennials rated messages with the thumbs up as more

passive-aggressive than Gen X did, and Gen X rated the thumbs up messages as more passive-aggressive than the Baby Boomers did. This supports anecdotal observations (e.g., Abril, 2022; Katz, 2022) that younger generations are reinterpreting the traditionally positive thumbs up and smile emojis as passive-aggressive. See Figure 7.
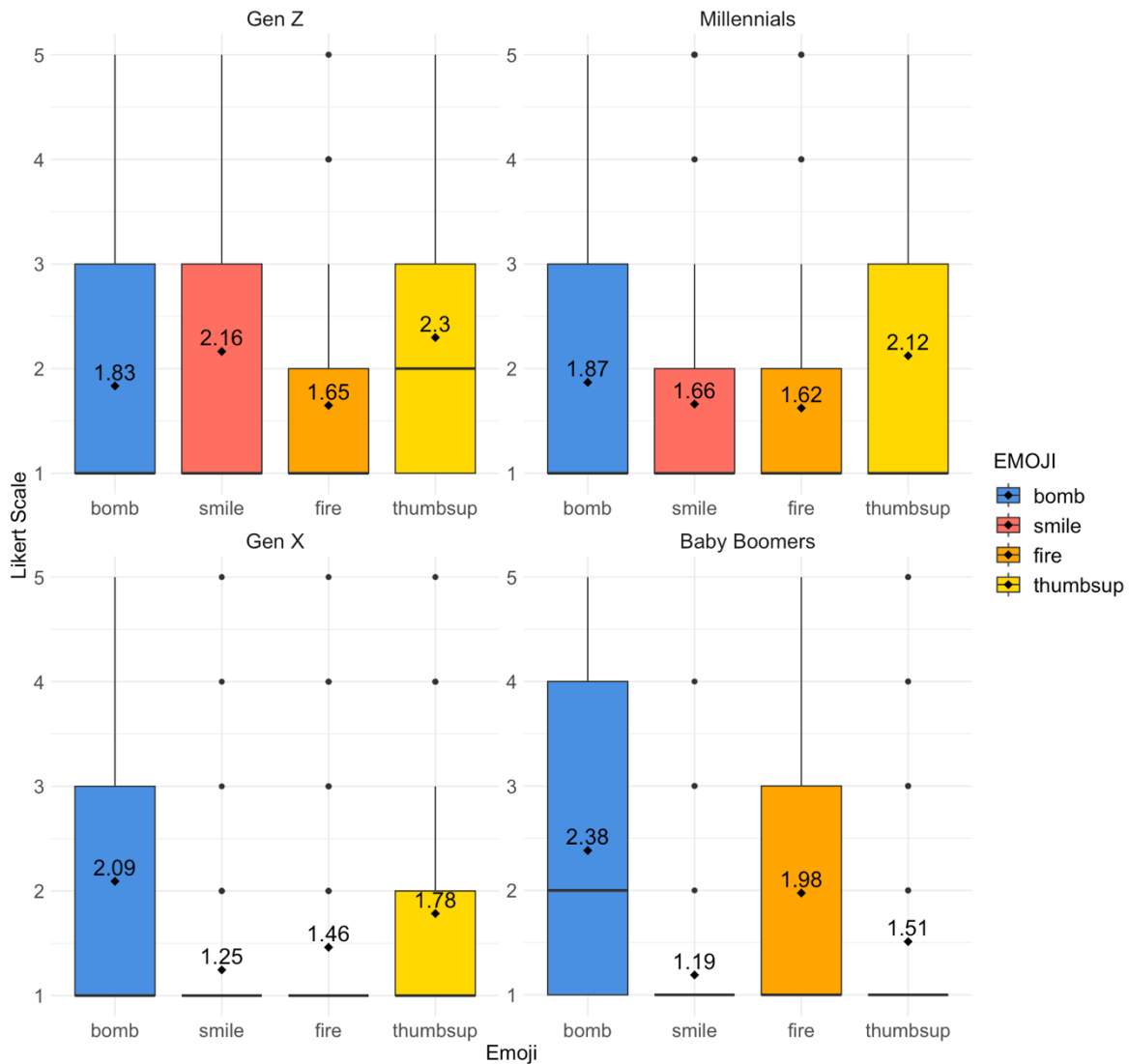

Figure 7. Likert Item for "Passive-Aggressive" by Generation and Emoji Type. The median is represented by the horizontal line, and the mean is represented by the black diamond in each box.

As regards "offensiveness" ratings, messages with the thumbs up emoji ($p < .05$) were perceived by Millennials as significantly more offensive compared to Gen Z respondents. Messages with the bomb emoji ($p < .05$) and with the fire emoji ($p < .01$) were perceived by Baby Boomers as significantly more offensive compared to the younger generations. See Figure 8.
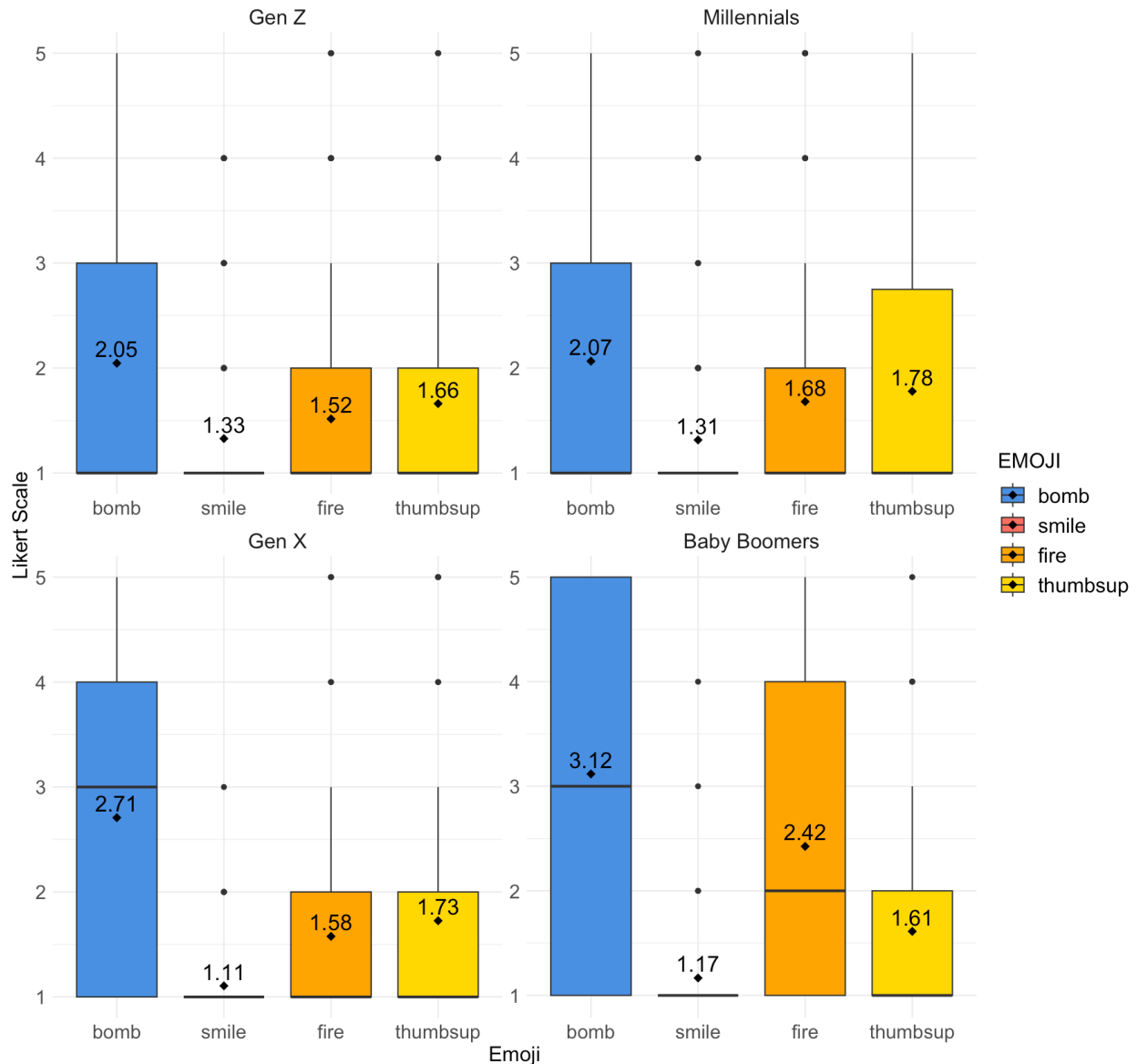
Figure 8. Likert Item for "Offensive" by Generation and Emoji Type. The median is represented by the horizontal line, and the mean is represented by the black diamond in each box.

Finally, compared to the younger generations, Baby Boomers rated messages with the fire emoji as significantly more threatening ($p < .01$) and messages with the thumbs up emoji ($p < .05$) as significantly less threatening, as shown in Figure 9. Although not a significant result, the figure also shows that some Gen Z respondents rated the smile emoji as threatening. These findings are evidence of generational differences in the perception of emoji toxicity.
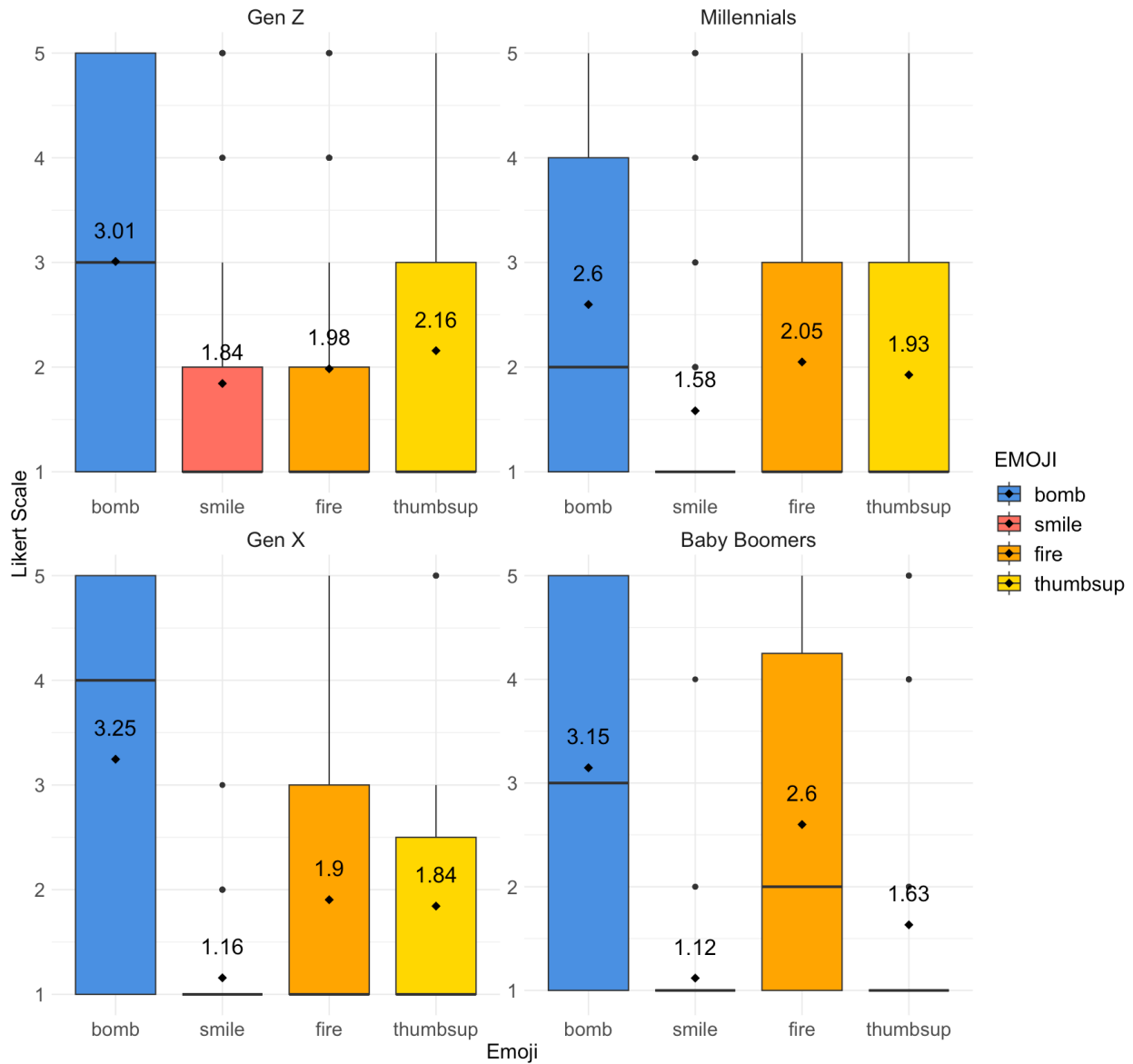
Figure 9. Likert Item for "Threatening" by Generation and Emoji Type. The median is represented by the horizontal line, and the mean is represented by the black diamond in each box.

Content Analysis

Thirty-two individuals, or 11.3% of all survey respondents, typed in a response to the open-ended response option for the question about where a message was most likely to appear. Some people chose the open-ended option for more than one message, resulting in a total of 50 responses. The gender and age distribution of the respondents is shown in Table 2.

GENDER/ GENERATION	BABY BOOMERS	GENX	MILLENNIALS	GENZ	Total
Female	4	4	7	2	17
Male	2	0	3	2	7
Non-binary	0	0	4	1	5
Prefer not to say	0	0	1	2	3
Total	6	4	15	7	32

Table 2. Gender and age distribution of survey participants who provided at least one open-ended response to the question: "In what context is a message like this most likely to appear?"

Five main themes were identified in the content analysis; their distribution is shown in Table 3.

Theme	Frequency	Percent of Total
Ambiguous	4	8%
Don't know/Not sure	12	24%
Infelicitous message	12	24%
Multiple contexts	12	24%
All contexts	4	8%
Other	6	12%
Total	50	100%

Table 3. Themes identified in the open-ended responses

A few individuals thought that the stimulus message was *ambiguous*. For example, a Gen Z respondent who chose not to indicate their gender explained that the message "OMG, you're going to regret that 😊" "[n]eeds more context. Could be 2 friends bantering (ironic, sarcastic sense) or a stranger threatening another." This comment is consistent with claims in the literature that the smile emoji could either be friendly or part of a threat.

The message: "You're going to regret that, fool 💣," which is triply negatively valenced (regret, fool, 💣), appeared to cause some confusion (*don't know/not sure*). One non-binary millennial commented that they were "[u]nsure where this would appear," and a Gen X female typed: "I hope nowhere but I'm not sure." The same message was thought to fit in *multiple contexts* by a female Baby Boomer, however. Another female Baby Boomer similarly opined that "[s]everal of the scenarios above are plausible" for the messages: "You're going to regret that, fool 😊" and "You're going to regret that 🔥."

A handful of responses commented that the stimulus message could be encountered in *all contexts*. One millennial female typed "all of them" for the messages "Meet me at the park at 6pm, fool 💣" and "OMG, you're going to regret that 😊." In both messages, the valence of the emoji conflicts with the valence of the text, so it is interesting that someone would think the messages could appear anywhere.

More revealing are comments expressing that the message was *infelicitous*, usually because the meanings of the text and the emoji were perceived to be incompatible. The fire and the thumbs up emojis, especially, were deemed infelicitous in relation to the message text.

1. *"🔥 usually means something really good and the text conflicts with that"* (female, Millennial, message: "You're going to regret that 🔥")

2. *"The bomb means something should be like, mind blowing, and the actual text doesn't match that."* (female, Millennial, message: "You're going to regret that 💣 ")

3. "This one is just weird. Hard to imagine a relationship sarcastic and ironic enough that this would make sense" (prefer not to say, Millennial, message: "Meet me at the park at 6pm, fool 👍")

The first and second comments imply that the 🔥 and 💣 emoji are positive and mitigate the threat of the textual message, while the third comment implies that the 👍 is sarcastic and ironic. All three comments were posted by Millennials.

The *other* code was assigned to comments that did not fit in any of the other categories. Most 'other' comments were made by males. Many of these mentioned more explicit details of the likely message context, such as the hypothetical person who sent the message, for example, a boss or partner.

4. *"private message, someone you don't know that well, neither friend nor stranger"* (male, Millennial; message: "Meet me at the park at 6pm 👍")

5. *"An official comment like from your boss or someone you just met and are trying to connect to, someone you have to be more formal with."* (male, Gen Z; message: "Great! Meet me at the park at 6pm 👍")

6. *"In a private message from an emotionally intimate friend/partner"* (non-binary, Gen Z; message: "Meet me at the park at 6pm 😊")

Some *other* comments were less straightforward to interpret, such as that of the male Baby Boomer whose comment on the message: "Meet me at the park at 6pm, fool 💣" was "*I must lead a sheltered life.*" The implication is that this older man has never seen such a message and that he finds it somehow shocking, possibly because it appears threatening.

Finally, a few comments addressed the threatening (or not) nature of the message directly. Two examples are given below.

> 7. *The emoji doesn't make sense with the sentence, it's not funny to diffuse the threat. Reminds me of when my mom would try to swear. Didn't go well.* (female, Gen X; message: "You're going to regret that 👍")
>
> 8. *Sounds like something that would happen in chat in a game, so that fudges the characteristics where "threatening" does not mean a real-life threat.* (male, Millennial; message: "You're going to regret that, fool 🔥")

These comments attempt to resolve what the respondents perceived as an inconsistency between the threatening textual message and the funny and playful emoji, attributing the first case to possible incompetence of the emoji user and casting the second as something that would happen not in "real life" but "in a game." In both cases, the emoji-containing messages are concluded to be not actually threatening, despite the threatening textual content.

In sum, the responses to the open-ended context option underscore that people rely on contextual clues such as the identity of the message sender and their relationship to the message receiver when interpreting messages that contain emojis. Further, the comments that specifically note that the emoji used did not match the content reveal the respondents' understanding of the emoji, as well as the perceived importance of congruence between emoji and text in message interpretation. Except for the comments categorized as *other*, the threatening 'regret' message generated more comments, suggesting that its use with an emoji was more unusual and required more explanation than the neutral 'park' messages.

Discussion

In this study, we asked to what extent different demographic groups (dis)agree on the interpretation of messages followed by an emoji, and how that varies according to the particular emoji, especially in uses that could be perceived as toxic. Significant differences were found in how demographic groups perceive emoji-based toxicity. In the following discussion, high ratings for passive aggressiveness, offensiveness, and threateningness are considered evidence of toxicity.

Non-binary respondents and those who preferred not to indicate their gender differed from the female and male survey respondents; they both more often rated the emoji-containing messages as more sarcastic, more passive-aggressive, and more offensive. These findings resonate with evidence that marginalized gender groups face increased online toxicity, priming them to perceive potential threats (Scheuerman et al. 2018). Additionally, our overall gender findings validate those of Herring and Dainas (2018, 2020), who found that female and male respondents to their survey did not differ from each other, but respondents who listed their gender as 'other' differed in their understandings of (non-toxic) emoji functions in Facebook comments.

These findings highlight the need for content moderation approaches that account for the perspectives and experiences of non-binary and gender diverse users.

Overall, self-identified females and males did not differ from one another in their ratings of emoji containing messages. Only Millennials showed marginally significant differences, with males rating messages as less offensive than females. That female Millennials rated emoji-containing messages as more offensive recalls Jones et al.'s (2020) finding that women rated negative face emoji more negatively than men did, due to an "emotional negativity bias" in facial emotion processing. It is not clear, however, why this pattern should be found only for Millennials. More data are needed, especially for older generations, which are less represented among our survey respondents, to determine whether this pattern is present in other generations.

Different generations perceived the emoji containing messages significantly differently. Overall, Gen Z respondents rated messages as more sarcastic or ironic compared to older generations. At the older end of the age spectrum, Baby Boomers rated the messages as more offensive. Herring and Dainas (2020) also found age-related differences in understanding of emoji functions in Facebook comments, with younger respondents providing more nuanced interpretations than older respondents. While younger users see clever irony in certain emoji uses, older adults may interpret those same emojis literally. This makes sense, given that Gen Z users grew up encountering emojis more often than other generations and use them more frequently (Adobe Fonts, 2022); thus, they may better grasp the nuances of emoji intent. This is supported by the survey responses to the question about frequency of emoji use in social media. Millennials and Gen Z more often than the older generations said they use emojis "All the time," while Baby Boomers and Gen X said they used emojis "Most of the time" or "Occasionally." A further factor is that young adults also use more sarcasm, and are better at identifying it, than older adults (Garcia et al., 2022; Rothermich et al., 2021).

Perceived emoji toxicity varied according to the emoji used, as well as the generation of the respondent. The messages containing the bomb emoji elicited more negative interpretations overall, as expected, while the messages containing the smile emoji received friendlier ratings. However, the results for messages with the fire and thumbs up emoji were more ambiguous. Younger users rated the fire emoji as friendly, meaning that it signifies something "hot" or "great." This suggests a shift in the meaning of the fire emoji from potentially representing anger to signaling excitement (i.e., from its literal to a metaphorical meaning). At the same time, the thumbs up emoji was associated with sarcastic meanings by Gen Z and Millennials, while Baby Boomers saw it as genuinely positive and playful. These results lend empirical support to the observations of Katz (2022) and Wagner and Yu (2021) that the thumbs up emoji is taking on ironic and negative connotations for younger users. We also found empirical evidence that younger users interpret the smile emoji as passive-aggressive and less playful than older generations, consistent with anecdotal observations in Abril (2022) and Dictionary.com (2022). Conversely, younger respondents interpreted the bomb emoji and the fire emoji as more playful than older respondents did, providing a partial answer to the question posed at the beginning of this article: Would Gen Z

interpret violent emoji such as bombs, fire, and guns ironically, and would they consider such uses non-threatening?

Table 4 summarizes the main differences found in emoji interpretation between the youngest generation, Gen Z, and the oldest generation, Baby Boomers, as well as the overall findings for each of the four emoji. These interpretations reflect patterns visualized in Figures 4-9, which show significant associations between emoji type and participant generation, independent of message content or context. Toxic interpretations are highlighted in boldface font in the table.

Emoji	Overall	Gen Z	Baby Boomers
😊	Most friendly of the emoji	**More passive-aggressive,** less playful	Less threatening, least passive-aggressive
👍	Most ambiguous of the emoji across generations	**More passive-aggressive,** less friendly	More friendly, more playful, less sarcastic, less threatening
🔥	Least sarcastic or ironic of the emoji	More friendly, more playful, **more passive-aggressive**	**More offensive, more threatening**
💣	**Most offensive and most threatening of the emoji**	More playful, less passive-aggressive	**More passive-aggressive, more offensive**

Table 4. Summary of generational differences (Gen Z vs. Baby Boomers) in emoji interpretation

Message content and valence influenced the perception that a message was threatening. When the message content was "You're going to regret that," respondents rated the message as significantly more toxic: more sarcastic, passive-aggressive, offensive, and threatening. They also had more trouble identifying contexts in which the message might appear, presumably because emoji use is not conventionally associated with toxicity. Messages with negative valence were also associated with higher levels of perceived offensiveness and threat. On one hand, these results confirm the expectations that we had when designing the survey. On the other hand, they provide statistically significant evidence that the potential toxicity of messages that contain emojis is dependent on their content and the context in which they appear.

More generally, the study's findings underscore the subjective and nuanced nature of emoji interpretations, particularly when used in potentially toxic messages. The significant differences across genders, generations, emoji types, and contexts highlight the need for greater awareness and contextual understanding in online communication and content moderation efforts.

Conclusion

To our knowledge, this is the first study of emoji understanding to consider demographic differences in perception of toxicity in emoji use. Our findings expand previous findings of gender and age differences in emoji interpretation (cf. Herring & Dainas, 2018, 2020) to potentially toxic

messages. Additionally, our analyses provide empirical evidence of the extent to which, and in what ways, the meanings of specific emojis are shifting across generations. The study thus underscores the importance of considering both toxic uses and demographic factors in the study of emoji understanding.

At the same time, although the literature reports that negative emojis can shift the meaning of neutral text and that positive emojis can amplify threatening text (Matulewska & Gwiazdowicz, 2020; Wagner & Yu, 2021), our efforts to reproduce these findings met with only partial success. Our neutral emoji-containing message (park) was primarily construed as friendly and playful, regardless of the emoji used, while the threatening message (regret) that contained an emoji was interpreted as playful as much or more than it was interpreted as passive-aggressive or offensive, and nearly as much as it was interpreted as threatening (see Figure 3). Moreover, sentences containing the regret message plus an emoji were often considered infelicitous or uninterpretable, as revealed through the content analysis. These findings are evidence of an overall positivity bias associated with emoji use (cf. Herring & Ge-Stadnyk, 2024). More unambiguously toxic stimuli and/or richer contextual set-ups might be needed to reproduce the online observations of Matulewska and Gwiazdowicz (2020) and Wagner and Yu (2021) in a controlled context. The survey methodology itself has inherent limits, as self-reported perceptions may not fully capture real-world reactions and interpretations.

Another limitation of this study is that it is situated within a U.S. English-speaking context, which can influence perceptions of offensiveness and toxicity. The United States' protections for free speech under the First Amendment (U.S. Const. amend. I) shape a cultural environment where the boundaries of acceptable speech, including hate speech, are relatively wide compared to many other countries like France and Germany. Thus, participants' interpretations of offensiveness and threat may reflect a U.S.-centric viewpoint. Relatedly, although we accounted for English proficiency in our screening criteria, we did not differentiate between native and non-native speakers in our analysis.

Other limitations include the relatively small numbers of non-binary individuals, males, and members of older generations who responded to our survey; the fact that the only demographic variables that were included in the analysis were age and gender; and that we only considered receiver demographics. Further, our analysis focused on a limited set of four common emojis, evaluated on five Likert scales, which almost certainly does not capture all possible user interpretations.

These limitations notwithstanding, our findings have broader implications, for example, for content moderation policies and legal considerations involving digital communication within the U.S. Failure to account for divergent interpretations of potentially toxic emoji usage across generations and genders could lead to bias in emoji meaning detection and moderation. To ensure safe and comfortable online interactions, more nuanced, contextually and demographically aware computational models are needed to facilitate emoji toxicity detection and content filtering tailored to different user groups (Kim et al., 2022; Kirk et al., 2021). In legal contexts, there is a need for

nuanced approaches that recognize the context and group-specific meanings of emoji, especially among youth, in cases where the interpretation of emoji usage is in question (Danesi, 2021; McMahon et al., 2019; Wagner & Yu, 2021). Improvement of content moderation and legal understanding of communication involving emojis are important and promising avenues for future research.

Other avenues for future research suggested by this study include demographic differences in emoji perception in other cultural contexts, as well as investigations involving a larger set of emojis, including potentially toxic ones like the water pistol and the knife. The line of research developed in this study could also be expanded to include demographic information about the sender, such as age and gender (cf. Cui, 2022). Equally important, additional receiver demographic variables such as race/ethnicity, religion, sexuality, and self-identification as transgender should be considered, since identities associated with these variables are frequent targets of online abuse (Casula et al., 2021; Kim et al., 2022; Scheuerman et al., 2018). Finally, we primarily focused on passive-aggression, offensiveness, and threateningness as measures of toxicity in this study, leaving room for future research to explore other facets of online toxicity, such as abuse and harassment, for a more comprehensive assessment of emoji perception in digital communication.

Notes

1. The name of this emoji has since been changed to 'enraged face.'

2. 'Threat' was the most common interpretation assigned to a sentence of this form in a study of Danish subjects (Bojsen-Møller, 2020). It was also sometimes interpreted literally, i.e., that the message producer was simply expressing that the addressee would come to regret their current actions or choices.

3. For purposes of analysis, we categorized the responses to these choices as public vs. private (context) and stranger vs. friend (familiarity).

4. Additionally, negatively valenced messages were rated highly for sarcasm/irony and playfulness. While some respondents perceived the messages ending with the epithet *fool* as offensive and threatening, as we expected, it appears that other respondents interpreted the negatively valenced messages as friendly banter.

References

Abril, D. (2022, December 12). Gen Z came to 'slay.' Their bosses don't know what that means. *Washington Post.* Retrieved September 10, 2023, from https://www.washingtonpost.com/technology/2022/12/12/gen-z-work-emojis/

Adobe Fonts. (2022). The future of creativity: 2022 U.S. emoji trend report: How Americans are using emoji. Adobe Blog. Retrieved September 10, 2023, from https://blog.adobe.com/en/publish/2022/09/13/emoji-trend-report-2022

Bojsen-Møller, M., Auken, S., Devitt, A. J., & Christensen, T. K. (2020). Illicit genres: The case of threatening communications. *Sakprosa*, *12*(1), 1-53.

Capobianco, M. (2022, November 13). I asked teenagers, 30-somethings, and people in their 50s what these emojis mean, and i got verrry different responses from each generation. *BuzzFeed.* Retrieved September 10, 2023, from https://www.buzzfeed.com/mollycapobianco/different-generations-emoji-use

Casula, P., Anupam, A., & Parvin, N. (2021, June). "We found no violation!": Twitter's violent threats policy and toxicity in online discourse. In *Proceedings of the 10th International Conference on Communities & Technologies-Wicked Problems in the Age of Tech* (pp. 151-159).

Cui, J. (2022). Respecting the old and loving the young: Emoji-based sarcasm interpretation between younger and older adults. *Frontiers in Psychology*, *13*, 897153.

Częstochowska, J., Gligorić, K., Peyrard, M., Mentha, Y., Bień, M., Grütter, A., West, R. (2022, May). On the context-free ambiguity of emoji. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 16, pp. 1388-1392).

Danesi, M. (2016). *The semiotics of emoji: The rise of visual language in the age of the internet*. Bloomsbury.

Danesi, M. (2021). The law and emojis: emoji forensics. *International Journal for the Semiotics of Law-Revue internationale de Sémiotique juridique*, *34*, 1117-1139.

Dictionary.com. (2022). How Gen Z uses emoji: A guide for millennials. *Dictionary.com.* Retrieved September 10, 2023, from https://www.dictionary.com/e/gen-z-explains-emoji-to-millennials/

Ferreira V. (2016, March 14). 'Meet me in the library 🔫 🔪 💣': Police and judges grapple with the meaning of 'offensive emoticons.' *National Post*. Retrieved November 19, 2023, from https://nationalpost.com/news/canada/meet-me-in-the-library-%F0%9F%94%AB-%F0%9F%94%AA-%F0%9F%92%A3-police-and-judges-grapple-with-the-meaning-of-offensive-emoticons

Garcia, C., Țurcan, A., Howman, H., & Filik, R. (2022). Emoji as a tool to aid the comprehension of written sarcasm: Evidence from younger and older adults. *Computers in Human Behavior*, *126*, 106971.

Goldman, E. (2022). 2021 Emoji law year-in-review. On *Technology & Marketing Law Blog.*

Herring, S. C., & Dainas, A. R. (2018, June). Receiver interpretations of emoji functions: A gender perspective. In *Proceedings of the 1st International Workshop on Emoji Understanding and Applications in Social Media* (*Emoji2018*).

Herring, S. C., & Dainas, A. R. (2020). Gender and age influences on interpretation of emoji functions. *ACM Transactions on Social Computing*, 3(2), 1-26.

Herring, S. C., & Ge-Stadnyk, J. (2024). Emoji and illocutionarity: Acting on, and acting as, language. In M. Gill, A. Malmivirta, & B. Warvik (Eds.), *Structures in discourse: Studies in interaction, adaptability, and pragmatic function*. John Benjamins.

Jane E. A. (2017). Feminist flight and fight responses to gendered cyberhate. In M. Segrave & L. Vitis (Eds.), *Gender, technology and violence* (pp. 45–61). Routledge.

Jones, L. L., Wurm, L. H., Norville, G. A., & Mullins, K. L. (2020). Sex differences in emoji use, familiarity, and valence. *Computers in Human Behavior*, *108*, 106305.

Katz, L. (2022, Oct. 19). Gen-Z, you'll have to pry the 'hostile' thumbs-up emoji from my old, dead hands. *Cnet.com.*

Kim, J., Wohn, D. Y., & Cha, M. (2022). Understanding and identifying the use of emotes in toxic chat on Twitch. *Online Social Networks and Media*, *27*, 100180.

Kirk, H. R., Vidgen, B., Röttger, P., Thrush, T., & Hale, S. A. (2021). Hatemoji: A test suite and adversarially-generated dataset for benchmarking and detecting emoji-based hate. *arXiv preprint arXiv:2108.05921.*

Matamoros Fernandez, A. (2018). Inciting anger through Facebook reactions in Belgium: The use of emoji and related vernacular expressions in racist discourse. *First Monday*, 23(9).

Matulewska, A., & Gwiazdowicz, D. J. (2020). Cyberbullying in Poland: A case study of aggressive messages with emojis targeted at the community of hunters in urbanized society. *Social Semiotics*, 30(3), 379-395.

McMahon, M. M., & Kirley, E. A. (2019). When cute becomes criminal: Emoji, threats, and online grooming. *Minnesota Journal of Law Science & Technology*, 21, 37.

Rothermich, K., Giorio, C., Falkins, S., Leonard, L., & Roberts, A. (2021). Nonliteral language processing across the lifespan. *Acta psychologica*, *212*, 103213.

Sanjay, S. (2020, October 1). Why does Gen Z use emojis so weirdly? *Vice*. Retrieved September 10, 2023, from https://www.vice.com/en/article/7kpngb/gen-z-young-people-use-emojis-differently-than-millennials

Scheuerman, M. K., Branham, S. M., & Hamidi, F. (2018). Safe spaces and safe places: Unpacking technology-mediated experiences of safety and harm with transgender people. In *Proceedings of the ACM on Human-computer Interaction*, 2(*CSCW*), 1-27.

Steinberg, B. (2022, October 26). Gen Z wants you to stop using this popular emoji. *New York Post*. Retrieved September 10, 2023, from https://nypost.com/2022/10/26/gen-z-has-hidden-meanings-for-emojis-beyond-rude-thumbs-up/

Subramanian, J., Sridharan, V., Shu, K., & Liu, H. (2019, July). Exploiting emojis for sarcasm detection. In *International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation* (pp. 70-80). Springer, Cham.

Yurieff, K. (2021, February 15). Sorry, millennials. The 😂 emoji isn't cool anymore. *CNN Business*. Updated 9:36 AM EST, Mon February 15, 2021. Retrieved September 10, 2023, from https://www.cnn.com/2021/02/14/tech/crying-laughing-emoji-gen-z/index.html

Vocativ (2015). The eggplant emoji is the next frontline of online harassment: The harmless purple eggplant emoji is being used as a symbol of sexual power and degradation. [No longer available.]

Völkel, S. T., Buschek, D., Pranjic, J., & Hussmann, H. (2019, October). Understanding emoji interpretation through user personality and message context. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 1-12).

Wagner, A., Marusek, S., & Yu, W. (2020). Emojis and law: Contextualized flexibility of meaning in cyber communication. *Social Semiotics*, *30*(3), 396-414.

Wagner, A., & Yu, W. (2021). Machiavellian apparatus of cyberbullying: Its triggers igniting fury with legal impacts. *International Journal for the Semiotics of Law–Revue internationale de Sémiotique juridique*, 34(4), 945-963.

Wiegand, M., & Ruppenhofer, J. (2021, April). Exploiting emojis for abusive language detection. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume* (pp. 369-380).

Appendix A

Stimuli

1. Meet me at the park at 6pm 💣
2. Great! Meet me at the park at 6pm 💣
3. Meet me at the park at 6pm, fool 💣
4. You're going to regret that 💣
5. OMG, you're going to regret that 💣
6. You're going to regret that, fool 💣
7. Meet me at the park at 6pm 🔥
8. Great! Meet me at the park at 6pm 🔥
9. Meet me at the park at 6pm, fool 🔥
10. You're going to regret that 🔥
11. OMG, you're going to regret that 🔥
12. You're going to regret that, fool 🔥
13. Meet me at the park at 6pm 😊
14. Great! Meet me at the park at 6pm 😊
15. Meet me at the park at 6pm, fool 😊
16. You're going to regret that 😊
17. OMG, you're going to regret that 😊
18. You're going to regret that, fool 😊
19. Meet me at the park at 6pm 👍
20. Great! Meet me at the park at 6pm 👍
21. Meet me at the park at 6pm, fool 👍
22. You're going to regret that 👍
23. OMG, you're going to regret that 👍
24. You're going to regret that, fool 👍

Fillers

1. Well, you are entitled to your opinion 🥺
2. Well, you are entitled to your opinion 👐

Attention Check

Please select the message followed by the Red Heart emoji to show that you are paying attention to this question.

1. Going to take the new puppy for a walk 😊
2. Going to take the new puppy for a walk 💕
3. Going to take the new puppy for a walk ❤️
4. Going to take the new puppy for a walk 😍

Appendix B

Variable	Estimate	Std. Error	z value	Pr(>|z|)
MESSAGE_regret	1.37	0.11	12.70	**<0.001**
EMOJI_smile	-0.53	0.14	-3.68	**<0.001**
EMOJI_fire	0.98	0.17	5.85	**<0.001**
EMOJI_thumbs_up	-0.39	0.15	-2.61	**<0.05**
GENDER_male	-0.17	0.19	-0.88	0.38
GENDER_non_binary	-0.07	0.21	-0.33	0.74
GENDER_prefer_not_to_say	0.14	0.26	0.52	0.60
GENERATION_millennials	0.09	0.24	0.38	0.71
GENERATION_gen_x	-0.15	0.28	-0.52	0.60
GENERATION_baby_boomers	-0.08	0.30	-0.25	0.80
VALENCE_positive	-0.30	0.19	-1.53	0.13
VALENCE_negative	0.97	0.33	2.93	**<0.05**

Table A. Model coefficients: CONTEXT ~ MESSAGE + EMOJI + GENDER + GENERATION + VALENCE + (1 | PARTICIPANT)

Variable	Estimate	Std. Error	z value	Pr(>|z|)
MESSAGE_regret	-1.01	0.09	-11.09	**<0.001**
EMOJI_smile	0.71	0.13	5.35	**<0.001**
EMOJI_fire	-1.18	0.16	-7.23	**<0.001**
EMOJI_thumbs_up	0.06	0.14	0.40	0.69
GENDER_male	-0.27	0.17	-1.58	0.11
GENDER_non_binary	0.21	0.19	1.10	0.27
GENDER_prefer_not_to_say	0.05	0.23	0.23	0.82
GENERATION_millennials	-0.18	0.22	-0.79	0.43
GENERATION_gen_x	-0.11	0.26	-0.42	0.68
GENERATION_baby_boomers	-0.41	0.29	-1.42	0.15
VALENCE_positive	0.08	0.14	0.55	0.58
VALENCE_negative	-0.42	0.20	-2.10	**<0.05**

Table B. Model coefficients: FAMILIARITY ~ MESSAGE + EMOJI + GENDER + GENERATION + VALENCE + (1 | PARTICIPANT)
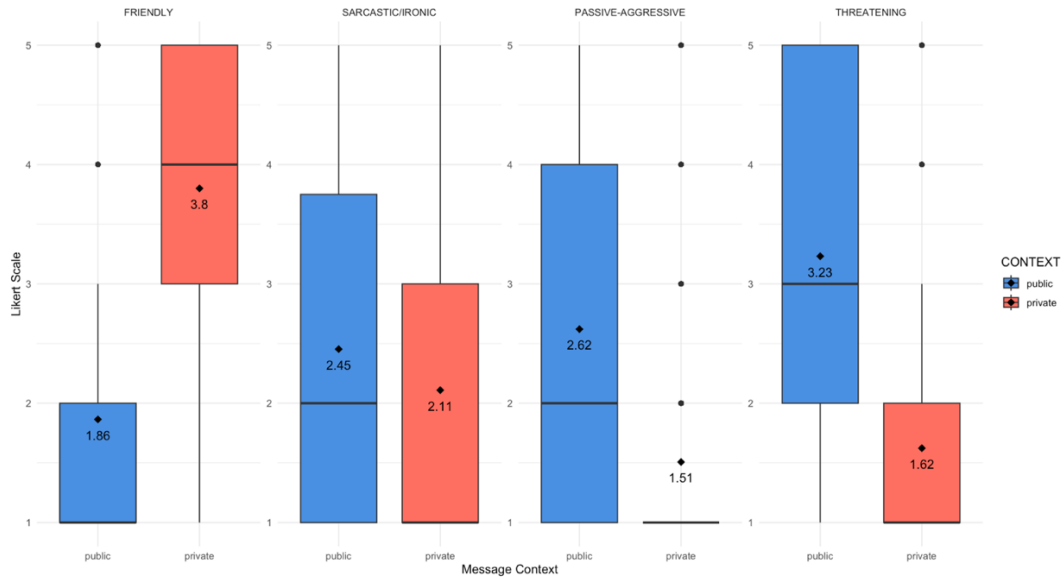
Figure A1. Likert Items by Perceived Message Context: Public vs. Private. The median is represented by the horizontal line, and the mean is represented by the black diamond in each box.
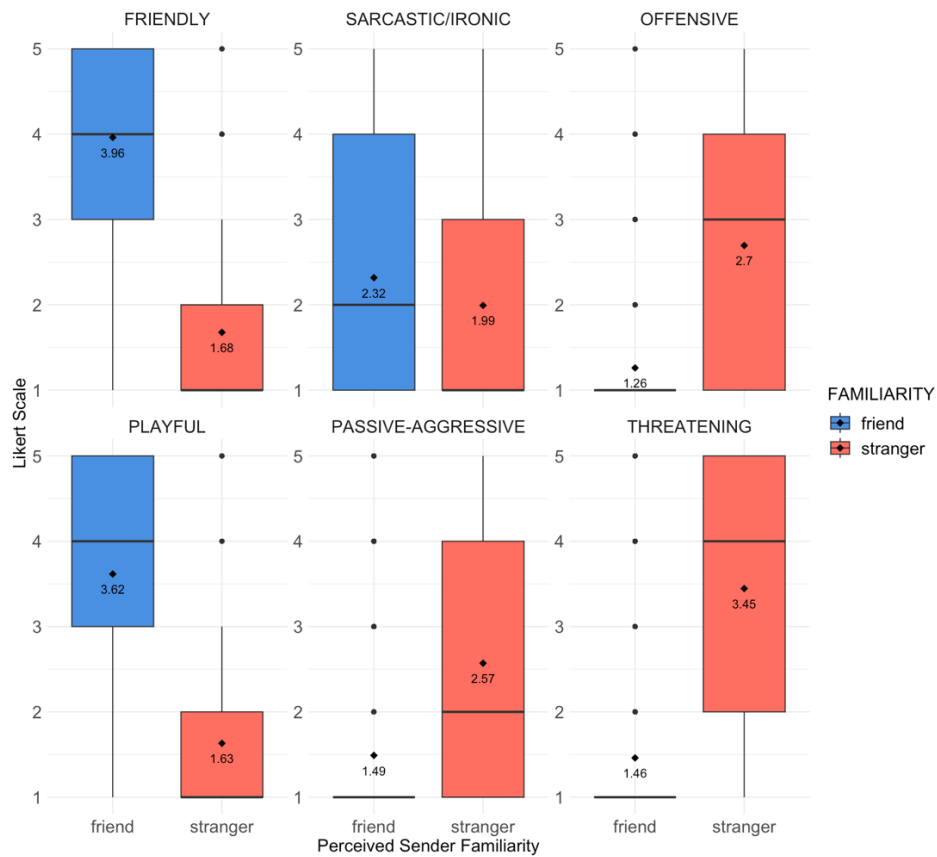

Figure A2. Likert Items by Perceived Sender Familiarity: Friend vs. Stranger. The median is represented by the horizontal line, and the mean is represented by the black diamond in each box.
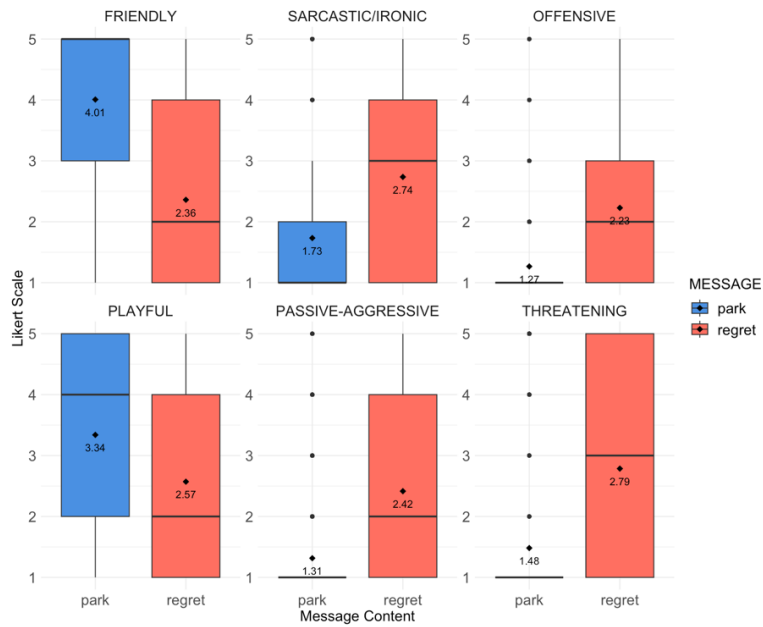
Figure A3. Likert Items by Message Content: Threatening vs. Non-threatening. The median is represented by the horizontal line, and the mean is represented by the black diamond in each box.
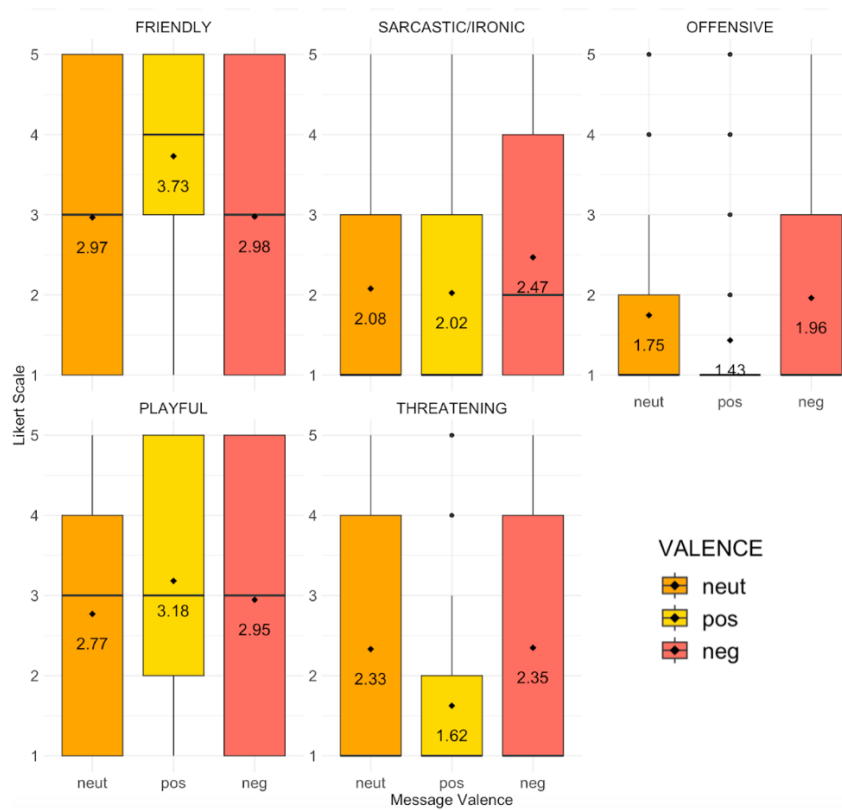

Figure A4. Likert Items by Message Valence: Neutral vs. Positive vs. Negative. The median is represented by the horizontal line, and the mean is represented by the black diamond in each box.