

Peaches and eggplants or... something else? The role of context in emoji interpretations

Benjamin Weissman*

Abstract. This paper presents the results of an experiment designed to measure interpretations of two emojis oft-discussed in popular culture, the eggplant and the peach. The experiment asked people to judge how sexual an emoji-containing text message was. The context surrounding these messages was manipulated across experimental conditions, altering both the preceding discourse and the presence of a sentence-final wink emoji. Unsurprisingly, the baseline interpretation of both the eggplant and peach emoji is euphemism. When one of these emojis is used in a context that strongly biases towards the non-euphemistic interpretation, ratings for sexualness decrease and variability increases. This suggests that participants are still able to access non-euphemistic interpretations of these emojis, but it must be under specific circumstances and will nonetheless come with a high degree of variability. Wink emojis added to messages containing non-euphemistic food emojis were also rated as more highly sexual (albeit still low on the rating scale), indicating an affective role for this emoji.

Keywords. emojis; euphemism; figurative language; pragmatics

1. Introduction. Much of the scholarly work on emoji ambiguity (e.g., Miller et al. (2016), Miller et al. (2017)) has rightfully focused on face emojis, which can each convey a wide range of potential meanings and sentiment, even in context. Outside of academia, among the most-discussed emojis in popular media are two non-face emojis – 🍆 and 🍑 – which themselves may carry the potential for ambiguity. These two in particular may carry the traditional food interpretations or euphemistic ones, initiated by visual resemblance and strengthened among emoji users through cultural transmission. The public has had a fascination with these emojis and their euphemisms, featured abundantly in blogs and articles; The 2018 Netflix movie “The Package” was even titled simply “🍆” for a time. It is evident that these emojis’ primary definitions are euphemistic ones, and their interpretations as produce are secondary if existent. In this study, I examine the effects of discourse context and the potentially-euphemism-signaling 😏 on interpretations of these two emojis.

2. Related Work. Linguistic research on emojis has explored various uses of emojis and emoticons, such as marking illocutionary force and communicating affect (e.g., Dresner & Herring (2010), Herring & Dainas (2017), Riordan (2017a)). Not only can emojis perform numerous communicative functions, a specific emoji can be ambiguous in its communicative contribution. Miller et al. (2017) found that interpretations of emojis vary even in context; nonetheless, pragmatics suggests that context must have an important role in emoji interpretation. As including discourse context and uptake in a corpus study may be arduous, other methods like conversation analysis (Gibson et al. (2018)), psycholinguistic experimentation (Cohn et al. (2018) Weissman & Tanner (2018)), and survey data (Miller et al. (2017)) can be of use in exploring contextual influences on interpretations.

*Author: Benjamin Weissman, University of Illinois at Urbana-Champaign (bpweiss2@illinois.edu)

One communicative role of non-face emojis can be to reduce ambiguity. Riordan (2017b) found that including emojis can disambiguate otherwise ambiguous words in sentences (e.g., “I got a ticket 🚗” vs. “I got a ticket 🚗”). But what if the emoji itself, as with the beloved peach and eggplant, is ambiguous? Such ambiguity poses a potential hurdle that projects (e.g., Barbieri et al. (2018), Barbieri et al. (2017), Wijeratne et al. (2017)) aiming to make emojis machine-readable will need to clear. Training computational models in the comprehension and production of emojis is an important goal at the nexus of computer science and linguistics; achieving human-like performance in both will lead to improvements in the state of the art of human-AI interactions and aid in the effectiveness of automated processes like online harassment detection.

The entire linguistic subfield of pragmatics focuses on interpretation of language in context; that surrounding discourse context can affect the interpretation of a given word or symbol is to be expected. Surrounding discourse, whether it be preceding or following, can and does influence the interpretation of any utterance or word. It stands to reason that these potentially ambiguous emojis could be partially, if not fully, disambiguated by surrounding context. This is especially helpful in probabilistic models - be they models for computers to disambiguate or representations of how humans disambiguate in real time (e.g., Jurafsky (1996)) - in which inputs from various sources will change the likelihood of one interpretation over another. Face emojis can be a key source of pragmatic context and often are used to guide the reader towards the intended interpretation of an utterance. Especially notable here is their use in communicating affect (e.g., Riordan (2017b) Riordan (2017a)). An affective 😊 that follows an otherwise ambiguous utterance may provide additional intentional information that helps the message receiver to disambiguate.

In this experiment, I investigate the role of preceding discourse context and following wink emojis on the interpretation of potentially ambiguous peach and eggplant emojis, as well as on two unambiguous food emojis without euphemistic counterparts. It is hypothesized that the euphemistic interpretation will be the baseline for the two critical emojis. When these emojis are used in contexts that bias strongly towards non-euphemistic food interpretations, it is hypothesized that the emojis will be interpreted non-euphemistically but with increased variability due to conflicting factors and the relative infrequency of encountering these emojis in such a context. It is also hypothesized that including a wink emoji after these emojis will lead to more euphemistic interpretations.

3. Method. To explore the role of discourse context and pragmatic marker face emojis, a survey methodology was utilized. In a norming study, participants (N = 60) were shown a series of individual food emojis with no surrounding discourse and asked to rate on a sliding scale how sexual each emoji is. The consensus highest-rated emojis were 🍆 and 🍑, while the consensus lowest-rated emojis were, sensibly, 🍷 and 🍵. These four were thus selected as the emojis to be used in the stimuli for this study.

Stimuli in this experiment were text message screenshots like those pictured in Figure 1. The message was described to be from a conversation between two best friends. The key utterance was always “I need some <emoji>” – a sentence frame that works with both the euphemistic and non-euphemistic interpretations of the emojis. There were two contextual ma-



Figure 1: Three example stimuli from the experiment. A - a euphemistic emoji in a euphemism-biasing context with a wink; B - a non-euphemistic emoji in a food-biasing context with no wink; C - a "mixed messages" example with a euphemistic emoji and wink in a food-biasing context

nipulations: preceding discourse (three levels) and wink (two levels). The preceding discourse could be bare (no preceding incoming text), a message that biases towards a non-euphemistic interpretation – “What are you getting at the grocery store?” – or a message that biases towards a euphemistic interpretation – “What are you doing this weekend?” A wink following the food emoji could be present or omitted.

In conjunction with the text message screenshot, participants (N = 968, recruited from Amazon Mechanical Turk) were asked to respond on a sliding scale to the question “How sexual is this message?” which was determined to be the most effective way to pinpoint the underlying concept. Participants were instructed to rate the (outgoing) message in blue. Screenshots were inserted into the surveys as images, removing the possibility for cross-platform visual differences of the emojis. A continuous, unnumbered sliding scale (as opposed to an ordinal integer scale) was used to allow for parametric analysis (e.g., Allen & Seaman (2007), Sullivan & Artino (2013), Weissman (2019)). Participants saw only one stimulus and provided one rating; this prohibits using a mixed-effects model for analysis but counteracts in-survey comparison bias in ratings. Following the rating screen, participants provided demographic information including age, gender, and overall emoji familiarity.

4. Results and Discussion. Ratings, collapsed into two emoji euphemism tendency groups (high: 🍌 and 🍎, low: 🍌 and 🍌) for the six relevant conditions are shown in Figure 2.

A linear model was run on the data, with the sliding scale rating (converted to 1-100) as the dependent variable and emoji euphemism tendency, preceding discourse context, and wink presence as predictors. There were significant main effects of all three predictors as well as significant interactions.

To better investigate the roles of discourse context and wink, two follow-up models were run, one on each of the emoji euphemism tendency groups. In the high euphemism tendency group, there were significant main effects of both context and wink presence as well as a significant interaction. η^2 values indicate that for the high euphemism group, discourse context ($\eta^2 = 0.22$) accounted for more variance than wink presence ($\eta^2 = 0.02$). Within the low euphemism group, there were significant main effects of both context and wink and no significant interaction. In this group, discourse context ($\eta^2 = 0.05$) and wink presence ($\eta^2 = 0.06$) account for around the same amount of variance.

Both preceding discourse context and the presence of a wink can aid in elucidating poten-

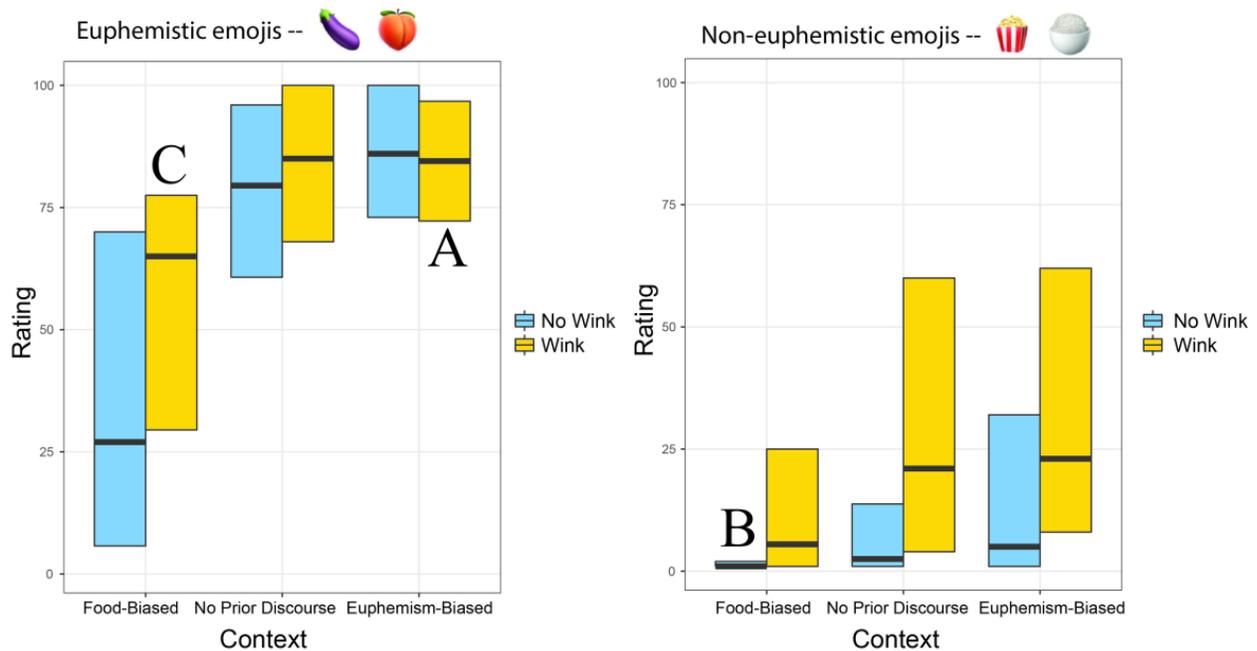


Figure 2: Boxplot of sliding scale rating results in six critical conditions. Grey line in each bar represents median value per condition. Without wink shown in blue on left; with wink shown in yellow on right. Conditions A, B, & C refer to the example stimuli from Figure 1.

tially ambiguous emojis. It is evident that in their baseline forms, with no prior discourse context, the eggplant and peach emojis are considered to be highly sexual in nature ($mean = 72.5$, $sd = 28.3$ without wink; $mean = 77.9$, $sd = 25.9$ with wink). When preceding discourse biases towards a euphemistic interpretation, there is a slight increase in rating and slight decrease in variance ($mean = 80.7$, $sd = 22.6$ without wink; $mean = 79.9$, $sd = 21.6$ with wink). If these emojis, strongly biased towards euphemistic interpretations, are placed in a context that strongly biases towards the non-euphemistic interpretation ($mean = 36.3$, $sd = 33.3$ without wink; $mean = 56$, $sd = 30.6$ with wink), variance increases significantly (apparent from visual observation and confirmed significant by Levene’s test). Such an increase in variance may be due to the fact that there are mixed signals in this message, or the fact that an occurrence like this is much less frequent and natural, or a combination of both factors.

Since the 🍿 and 🥚 carry no euphemistic interpretations (no obvious or familiar ones, at least), it cannot be said that discourse context and the presence of a wink disambiguate the emoji. Nonetheless, t-tests indicate that the inclusion of a wink emoji significantly increases the sexualness rating in all three context conditions. Riordan’s Riordan (2017b) Riordan (2017a) exploration of emoji-based affect marking can explain this finding; including a wink emoji, no matter how benign the preceding discourse, can introduce an element of playfulness or flirtation. It is, however, worth noting that despite the significant difference in ratings, the mean ratings for messages with these emojis are all under 30 on the 100-point scale - decidedly non-sexual compared to any messages with the euphemistic emojis.

Within the high euphemism emojis, there were two significant correlations with demographic factors. There was a significant positive correlation between rating and self-reported overall emoji familiarity ($r = .22, p < .001$) and a significant negative correlation between rating and age ($r = -.24, p < .001$). There was no significant effect of gender.

5. Conclusions. This work comprises the first to investigate interpretations of the notoriously euphemistic 🍑 and 🍆 emojis in different contexts. The euphemistic baselines of those emojis is confirmed. Results from the food-biased context interpretations indicate that while these emojis still carry the potential to be interpreted as their non-euphemistic meanings, they still carry a hint of salaciousness as compared to pruder food emoji counterparts like 🍷 and 🍿. Preceding discourse context and the presence of a wink emoji contributed to interpretations of the utterance as a whole, and it is noted that these factors could be utilized alongside corpus-based frequencies to aid in NLP efforts of emoji disambiguation. It may be especially valuable to include other emojis in surrounding contexts to train computational models of meaning-making.

Future work on these emojis could explore psycholinguistic processing components of the emojis, with a look at how our prior frequencies and expectations for sense disambiguation interact with contextual factors to influence on-line processing and meaning-making. Future work could also expand the notion of context to include interpersonal relationships as well. Interpretations of an eggplant emoji may differ significantly if a message comes from your best friend as opposed to, say, your grandmother. More broadly, it may prove valuable to investigate how speaker-specific information affects emoji usage and interpretation in a variety of contexts.

References

- Allen, I. Elaine & Christopher A Seaman. 2007. Likert scales and data analyses. *Quality Progress* 40. 64.
- Barbieri, Francesco, Miguel Ballesteros, Francesco Ronzano & Horacio Saggion. 2018. Multimodal emoji prediction. <https://arxiv.org/abs/1803.02392>.
- Barbieri, Francesco, Miguel Ballesteros & Horacio Saggion. 2017. Are emojis predictable? <https://arxiv.org/abs/1702.07285>.
- Cohn, Neil, Tim Roijackers, Robin Schaap & Jan Engelen. 2018. Are emoji a poor substitute for words? Sentence processing with emoji substitutions. *Proceedings of the 40th Annual Conference of the Cognitive Science Society*. 1524–1529.
- Dresner, Eli & Susan C. Herring. 2010. Functions of the nonverbal in CMC: Emoticons and illocutionary force. *Communication Theory* 20(3). 249–268. <https://doi.org/10.1111/j.1468-2885.2010.01362.x>.
- Gibson, Will, Pingping Huang & Qianyun Yu. 2018. Emoji and communicative action: The semiotics, sequence and gestural actions of face covering hand. *Discourse, Context and Media* 26. 91–99. <https://doi.org/10.1016/j.dcm.2018.05.005>.
- Herring, Susan C & Ashley Dainas. 2017. Nice picture comment! Graphics in Facebook comment threads. *Proceedings of the Hawai'i International Conference on System*

- Sciences (HICSS)* 50. 2185–2194.
https://aisel.aisnet.org/hicss-50/dsm/persistent_conversation/4/.
- Jurafsky, Daniel. 1996. A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive science* 20(2). 137–194.
- Miller, Hannah, Daniel Kluver, Jacob Thebault-Spieker, Loren Terveen & Brent Hecht. 2017. Understanding emoji ambiguity in context: The role of text in emoji-related miscommunication. *Proceedings of the Eleventh International AAAI Conference on Web and Social Media*. 152–161. <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15703>.
- Miller, Hannah, Jacob Thebault-Spieker, Shuo Chang, Isaac Johnson, Loren Terveen & Brent Hecht. 2016. “Blissfully happy” or “ready to fight”: Varying interpretations of emoji. *International AAAI Conference on Web and Social Media (ICWSM)*. 259–268.
- Riordan, Monica A. 2017a. Emojis as Tools for Emotion Work: Communicating Affect in Text Messages. *Journal of Language and Social Psychology* 36(5). 549–567. <https://doi.org/10.1177/0261927X17704238>.
- Riordan, Monica A. 2017b. The communicative role of non-face emojis: Affect and disambiguation. *Computers in Human Behavior* 76. 75–86. <https://doi.org/10.1016/j.chb.2017.07.009>.
- Sullivan, Gail M. & Anthony R. Artino. 2013. Analyzing and interpreting data from Likert-type scales. *Journal of Graduate Medical Education* 5(4). 541–542. <https://doi.org/10.4300/JGME-5-4-18>.
- Weissman, Benjamin. 2019. *The role of linguistic meaning in the concept of lying*: University of Illinois at Urbana-Champaign Dissertation.
- Weissman, Benjamin & Darren Tanner. 2018. A strong wink between verbal and emoji-based irony: How the brain processes ironic emojis during language comprehension. *PLoS ONE* 13(8). <https://doi.org/10.1371/journal.pone.0201727>.
- Wijeratne, Sanjaya, Lakshika Balasuriya, Amit Sheth & Derek Doran. 2017. A semantics-based measure of emoji similarity. *Proceedings of the International Conference on Web Intelligence* 17. 646–653. <https://doi.org/10.1145/3106426.3106490>.