Compositional Affordances of Emoji Sequences

Benjamin Weissman
Rensselaer Polytechnic Institute

Jan Engelen
Lena Thamsen
Neil Cohn
Tilburg University

Abstract

Emoji have become ubiquitous in digital communication, and while research has explored how emoji communicate meaning, relatively little work has investigated the affordances of such meaning-making processes. We here investigate the constraints of emoji by testing participant preferences for emoji combinations, comparing linearly sequenced, “language-like” emoji strings to more “picture-like” analog representations of the same two emoji. Participants deemed the picture-like combinations more comprehensible and were faster to respond to them compared to the sequential emoji strings. This suggests that while in-line sequences of emoji are on the whole interpretable, combining them in a linear, side-by-side, word-like way may be relatively unnatural for the combinatorial affordances of the graphic modality.

Introduction and Background

Emoji have become a ubiquitous part of digital communication, and research on emoji has primarily focused on the relationship of emoji to language. This linguistic focus likely arises because emoji are integrated with keyboards to be used together with text, and as such maintain linear sequencing like writing. Yet, this linguistic focus has not considered whether such constraints also limit emoji communication, and work has yet to examine how this affects their pictorial affordances. Here we highlight these aspects of emoji comprehension, focusing on emoji-only sequences.

Within the psycholinguistics language processing literature, a majority of studies on emoji have focused on emoji occurring alongside language (e.g., Barach et al., 2021; Beyersmann et al., 2023; Paggio & Tse, 2022; Scheffler et al., 2022; Weissman & Tanner, 2018). Some processing and comprehension research has assessed standalone singular emoji as well, either as part of a linguistic exchange (Holtgraves & Robinson, 2020) or not (e.g., Homann et al., 2022; Kaye, Darker et al., 2022; Kaye, Rodriguez-Cuadrado et al., 2021; Weissman et al., 2023). This research has provided fascinating insights into the time course and resources involved in processing emoji in a variety of contexts. A gap in the emoji processing literature exists, however, when it comes to the processing of multi-emoji expressions.

The primary theoretical debates about multi-emoji sequencing address the degree to which these sequences display properties similar to grammars in written languages. For example, recent analyses of multi-emoji (excluding instances of the same emoji repeated) sequences on the Chinese social media platform Sina Weibo have argued that emoji sequences exhibit word order patterns (Ge & Herring, 2018; Herring & Ge, 2020). Herring and Ge (2020) were able to assign subject, object, and verb labels (in differing orders) to 83% of a set of 300 emoji sequences; in their set, the predominant basic word order (appearing in roughly 60% of the labeled examples) was OVS. They view emoji as an "emergent graphical language" (Ge & Herring, 2018) with both a vocabulary and a preliminary grammar, supported by the evidence that emoji sequences exhibit a sentence-like linear order at a high rate.

A different view on emoji sequencing has come from Gawne and McCulloch (2019), who analyzed a corpus of over a billion emoji uses. Overall, utterance length of emoji-only sequences followed a Zipfian-like decay (Zipf, 1935), with single-emoji expressions produced roughly twice as frequently as two-emoji sequences, which in turn were more frequent than three or more emoji sequences. Analysis of the top bigram, trigram, and quadrigram combinations of emoji revealed that the majority of combinations featured repetition (e.g., 😍😍😍) at a much higher rate than words (McCulloch & Gawne, 2019). Since repetitions are quite rare in a similar analysis of words, they concluded that emoji may be better treated like co-speech gestures, which do feature this type of repetition. Additional analyses, one theoretical (Grosz et al., 2021) and one experimental (Pasternak & Tieu, 2022), support this idea that emoji pattern like co-speech gestures. Such similarities between text-emoji and co-speech gestures may arise because they share structural alignment as multimodal interactions where one modality has a grammar and another does not, all linked within a common cognitive architecture (Cohn & Schilperoord, 2022).

Other research has further found that full sequences comprised of only emoji remain relatively simplistic and limited in their productive capacities (Cohn et al., 2019). One of the only published experimental studies investigating the combinatorial properties of emoji on their own, Cohn et al. (2019) devised a production task where participants communicated in sequences of only emoji. The combinatorial nature of the produced emoji-only sequences remained simplistic, suggesting that emoji themselves do not have the properties that lend well to grammatically complex expressions (i.e., categorical roles, word orders, phrase structures). Emoji-only sequences, in this experiment, were required to be strung together linearly like words and thus were forced, rather unnaturally, into the categorical roles of language (e.g., noun and verb).

Some researchers have explored the extent to which machine-translation of English word sequences to emoji sequences is possible and useful (Wicke & Cunha, 2020). Despite progress in this endeavor, these authors readily admit such outputs are unnatural and not easily interpretable. Others, however, are more optimistic on the potential outcomes of text-to-pictogram machine translation efforts (e.g., Norré et al., 2021; Sevens et al., 2017; Vandeghinste et al., 2017), suggesting that pictogram sequences may be useful in navigating doctor-patient interactions or for individuals with intellectual disabilities for whom text-based online communication may be more difficult. Elsewhere, researchers have progressed in the creation of an iconic language that is intended to fully replicate linguistic structure and representational capacity in a fully graphic modality (e.g., Meloni et al., 2022).

These issues in the sequencing of simple images like emoji have been reflected recently in work by Morin (2023), who asks why pictorial ideographic systems have not emerged as generalized, self-sufficient communication systems. He argues that the graphic modality itself is limited in that it does not standardize in the way that the spoken modality of language inherently does. As a result, ideographic sequences are limited by their capacity to become grammaticalized with the same degree of sophistication as spoken/written languages.

In their response to Morin's argument, Cohn and Schilperoord (2023) agree that such ideographies are indeed limited, albeit for entirely different reasons. They counter that ideographies are altogether incapable of blossoming into the robust types of communication sketched by Morin, and by extension to fully grammatical emoji sequences. Cohn and Schilperoord provide extensive evidence that the graphic modality does allow for standardization, and that narrative graphic sequences (such as comics) display the same types of complex grammars as the syntax of spoken or signed languages. However, the real problem with ideographies is that they force the information structure and manner of presentation of the spoken modality onto the graphic modality. This mismatch arises between the graphic modality's expression of complex analog information in each unit and the spoken modality's conceptually simpler and linearly presented lexical items. Attempts to coerce the graphic modality into simple units with a linear sequence will be structurally (and thus cognitively) unnatural.

Though research has begun to acknowledge that graphic and other visual information shares underlying cognitive structures with language (e.g., Cohn, 2016; Cohn & Schilperoord, 2022; Holler & Levinson, 2019), this work has also acknowledged that pictures have affordances that differ from those of spoken or written languages. For example, pictures allow semantic compositionality in spatial relationships that are not strictly linear. Many uses of emoji combine within their linear constraints to create larger scenes. A recent example is the frequent collocation of the dotted-line face emoji 🫥 between trees 🌳🫥🌳, recreating the "Homer Simpson shrinking into the bushes" meme using emoji (Wright et al., 2023). Such cases demonstrate the creation of analog, spatial representations within the constraints of a forced linear side-by-side presentation.

An example of modality-specific combinatorial properties comes from Cohn, Murthy, and Foulsham (2016) and Cohn and Foulsham (2022), who investigated "upfixes," visual affixes of elements that appear above a character's head (e.g., a lightbulb or gears), typically in comics andcartoons.. These elements are understood not just by their content (i.e., whether the face and upfix had compatible meanings) but also by the placement of the upfix. Faces with upfixes positioned to the side of the face, instead of their canonical "floating above" placement, were rated as less comprehensible, suggesting specific combinatorial constraints on how elements should interact. Nothing about the face or upfix inherently changes because of a side-by-side representation, but that layout does not follow the conventional spatial constraints of comics; this notion is reinforced by the finding that participants' comic-reading expertise modulates this effect. This research thus provides evidence for modality-specific preferences for how elements are combined.

Morin argues that emoji are "not yet ready to replace writing" (2023, p. 15), due primarily to the lack of standardization, but he suggests they may someday be able to if, as his account predicts, standardization of emoji continues and becomes more generalist over time. Feldman's (2023) response to Morin, along with recent experimental evidence (Weissman et al., 2023), suggests that standardization of emoji is already underway. Whether a robust, self-sufficient ideography will someday emerge from future iterations of a standardized emoji set remains a topic of debate, but the present study explores whether such possibilities may be limited by the affordances of emoji from the start.

These debates about emoji sequencing and language notwithstanding, no research has yet directly questioned the naturalness of linear pictographic sequencing in the first place. Emoji presented in a language-like linear sequence result in emoji expressions that often are inexplicit about their relationships, leading to considerable inferencing on the part of the reader to arrive at a full-fledged proposition. In spoken language, grammar is typically thought to mediate this inferencing, providing a standard and formal system for structuring the relations between units (e.g., Chomsky, 1965; Culicover & Jackendoff, 2005; Langacker, 1987). Emoji, lacking a language-like grammar, seem to be considerably less explicit about how the units relate to each other (Cohn et al., 2019).

Here, we aim to examine further the issue of the pictorial affordances of emoji: To what extent does the technologically mandated linearity of emoji insertion demonstrate a capacity for combinatorial meaning-making? We compare linear emoji sequences with more picture-like analog depictions of the same emoji combinations, with the former requiring individual units presented side by side and the latter having more freedom for placement and relative sizing. Both depictions are compositional in nature but differ in the spatial constraints on such compositionality. Henceforth, "linear" refers to how emoji must typically be typed into a keyboard, featuring uniform sizing and language-like one after the other placement; "analog" comes from the visual language literature (Cohn, 2018) and refers to the relatively spatially unconstrained compositionality like that observed in drawing.

In the experiment reported in this article, the stimuli consist of multiword text expressions and corresponding two-emoji combinations. The experiment features two manipulations: congruence between text and emoji and presentation type of the two-emoji expressions. Emoji either match or mismatch the text phrase, and they appear either in a linear order like in a sentence or as a more spatially-pictorial analog depiction.

Through this comparison of presentation types, we examine the naturalness of emoji compositionality. If emoji compositionality respects the spatial linearity demonstrated by written languages (e.g., Danesi, 2016), we would expect to observe no processing cost for a linear presentation compared to an analog one. This result would suggest emoji can map directly onto their corresponding linguistic representations in a natural way. In this case, linear strings of compact conceptual information in the visual form, as presented by emoji, are just as natural as analog pictures, which are known to evoke iconic compositionality corresponding to scene perception (Võ, 2021). Contrarily, the presence of a processing cost to a linear emoji sequence may suggest that this manner of emoji combination is less natural than the analog format and that this graphic modality affords a different type of compositionality.

Methodology

Stimuli

For this experiment, 20 stimuli were designed. This included three compounds (e.g., "dumpster fire," "apple tree") and 17 action phrases (e.g., "The man eats pizza," "The girl did yoga at the beach"), all in English. These stimuli explicitly aim to examine circumstances where emoji could combine to create meaning, either as a side-by-side linear sequence or graphically composed into a single analog scene. The classic yellow facial expression emoji were omitted from this experiment: Combining face emoji into a single representation would involve either putting that head onto a body (which might be strange) or incorporating the facial expression into the face of an otherwise non-emoji body, which would lose the sensation of it originating from the emoji. Instead, stimuli primarily utilized full-body emoji, realistic human face emoji, and object emoji.

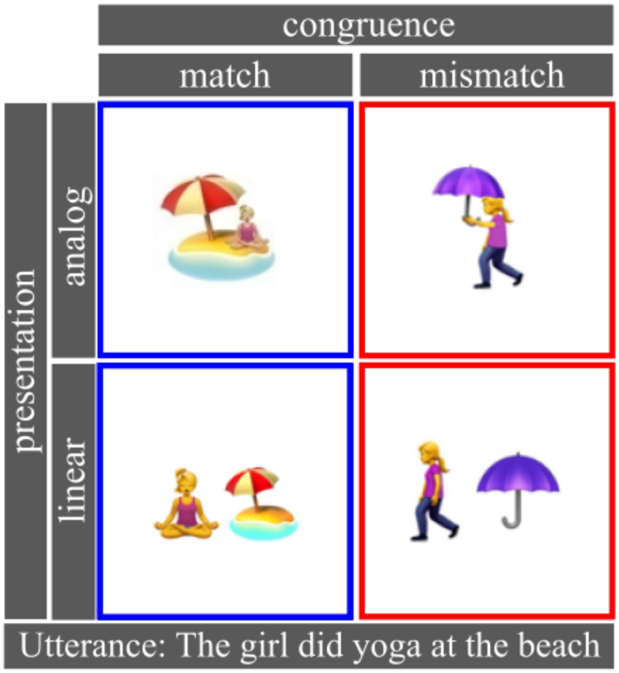


Figure 1. The four conditions used for stimuli. The top row depicts the analog presentation and the bottom row the linear presentation. The left column presents the match condition, in which the emoji expression matches the sentence; the right column presents a mismatch condition for this sentence.

For each item, two emoji depictions were created – one that presented the phrase in linear sentence order, as in the bottom row in Figure 1, and another that presented the phrase in an analog presentation, as in the top row in Figure 1. The linear presentation came from the two relevant emoji placed side by side without any resizing or alteration, exactly as they would appear in typical messaging apps. To achieve the analog presentation, the emoji images were resized, rotated, and/or overlapped. This essentially achieves a unified pictorial representation of the content referenced by the utterance, made from the same two emoji as the linear presentation condition. The full set of stimuli appears in the supplementary materials.

For each trial, the phrase could be either matching or mismatching with the emoji. For the mismatch condition trials, a phrase that matched another item in the experiment was used. These stimuli were counterbalanced in a 2x2 Latin square design with text-emoji congruence (match, mismatch) and presentation type (linear, analog) as the two conditions. Each participant thus encountered 20 critical stimuli throughout the experiment, with five items in each of the Congruence/Presentation combination conditions. The critical stimuli were interspersed with items from another experiment, which act as fillers here. These filler stimuli also probed a match/mismatch decision but featured single-word and single-emoji presentations.

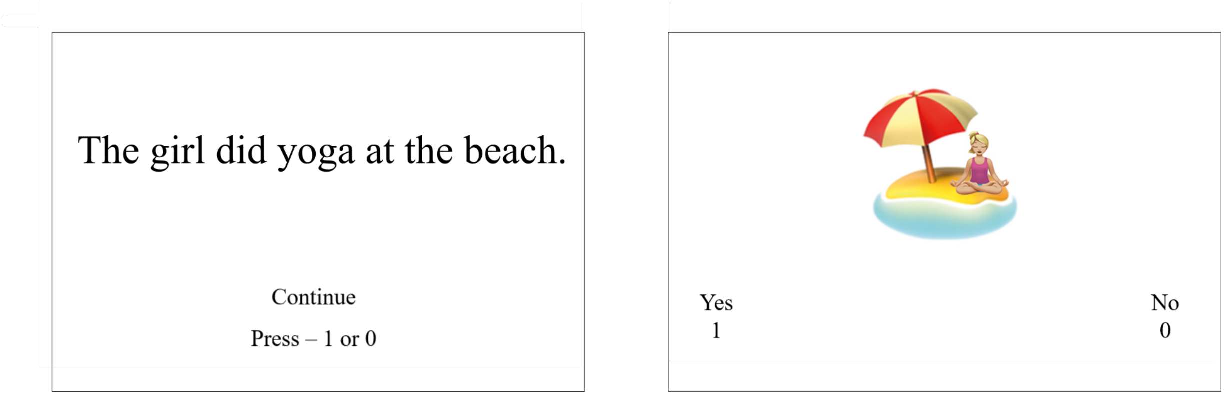
Participants

A total of 156 participants completed the experiment. Participants were required to be fluent, though not necessarily native, speakers of English, as assessed via self-report. Participants provided informed written consent, as approved by the Tilburg University Ethics Review Board. Three participants demonstrated patterns of consistent inattention, two of whom repeatedly responded in under 300 ms and one of whom repeatedly responded in over 20 seconds. These three participants were removed from the dataset; the final set thus includes responses from 153 participants (average age = 29.3 (SD = 8.85); 62 male, 85 female, 6 other).

Alongside a basic demographic questionnaire, participants completed an Emoji Language Fluency questionnaire created to gauge each individual’s experience and familiarity with emoji in general. Emoji Language Fluency did not correlate significantly with accuracy or response times in the experiment, nor did any of its questions; the survey itself appears in the supplementary materials. This questionnaire does indicate that participants were overall fairly familiar with emoji (average self-rated “emoji expertise” score = 4.8 (SD = 1.06) on a 1-7 scale).

Procedure

A match/mismatch response time experiment was utilized to explore processing time for the two stimulus types. Each trial in the experiment consisted of two screens. The first screen was untimed and presented a text-only phrase. When ready, participants pressed a button to move to the second screen, which presented the emoji combination in either its analog or linear form. On this timed screen, participants indicated whether the presented emoji combination matched or mismatched the phrase on the first screen by pressing the corresponding button on the keyboard. The second screen’s being timed provides a measurement of processing time of the emoji depiction, which allows for comparison between analog- and linear-presentation items. A sample trial is depicted in Figure 2. Participants accessed the experiment in a Qualtrics-hosted online survey where a response time experiment was presented using the jspsych plugin (De Leeuw, 2015).



The girl did yoga at the beach.

Figure 2. Example of experimental trial. First, participants read the text on Screen 1, then proceeded to Screen 2, the timed response (YES/NO) to whether the depiction matched the text from Screen 1.

Data Analysis

Responses under 300 ms (n=2) or over 2.5 standard deviations from the grand mean (n=70) were tagged as outliers and removed from the dataset. In total, 2.3% of data points collected were removed. Mixed-effects models were run using the *lme4* package in R (Bates et al., 2015) to account for item- and participant- level variance. Models including random slopes did not converge, so the models in the present analysis include random intercepts only. Sum coding was utilized for all models (see Brehm & Alday [2022] for more).

To examine participants' responses to the text-emoji congruence, a logistic mixed effects model with sum coding was run on the outcome variable of binary response data (0 – inaccurate response, 1 – accurate response), with text-emoji congruence (match/mismatch), presentation (analog/ linear), and the interaction as predictors as fixed effects and participant and item (the emoji displayed) both as random intercepts.

Response times for these judgements were similarly analyzed using a linear mixed effects model with sum coding, taking response time (measured in ms) as the outcome variable, again with congruence (match/mismatch) and presentation type (analog/linear) and the interaction as predictors, and participant and item (the emoji displayed) as random effects. All non-outlier responses were analyzed, including incorrect responses, as we assume, especially in the Congruent condition, that participants were still engaging in meaningful meaning-making processes.

Results

Figure 3 shows the response data, sorted by text-emoji congruence and color-coded by presentation type. As expected, matching emoji were deemed more congruent than mismatching emoji. Accuracy was extremely high for both incongruent conditions regardless of presentation type: A "no" response was correctly provided on 98.5% of analog trials and 99.1% of linear trials (*p* = 0.23). Presentation type did, however, significantly vary responses for matching text-emoji items. Here, analog emoji depictions were rated as congruent (97.2%) with the preceding text significantly more often than linear depictions (87.1%), (*p* < .001). This led to a significant interaction between congruence and presentation type (z = 2.05, *p* = 0.04).

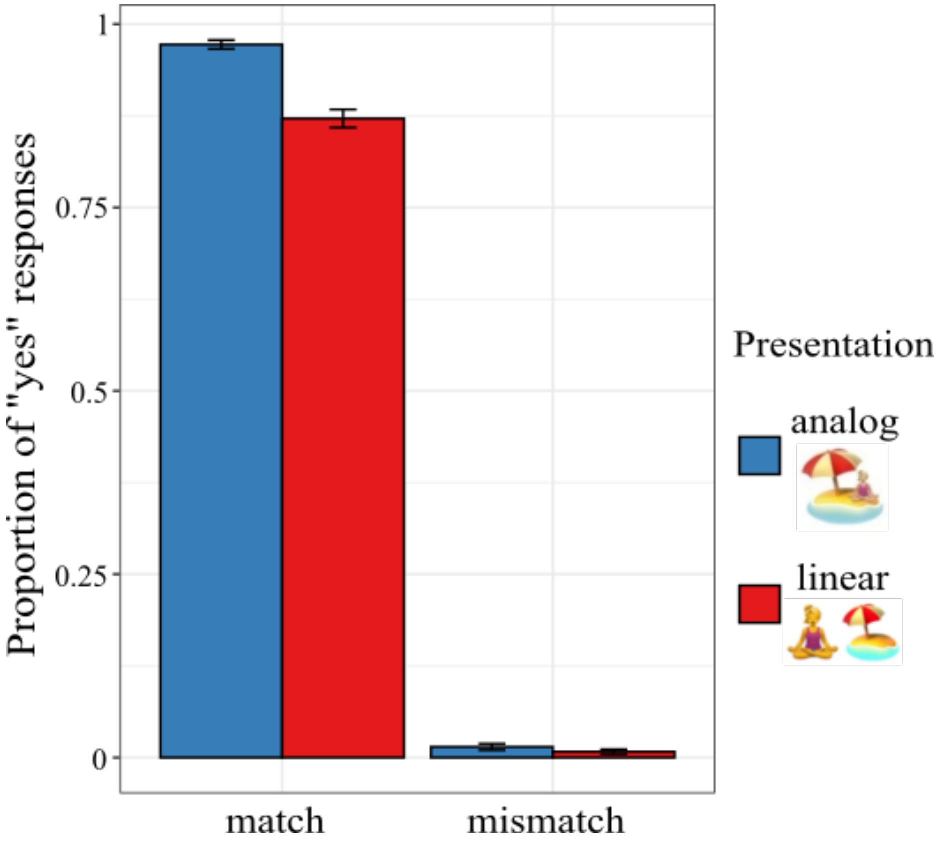


Figure 3. Proportion of "yes" responses for the emoji expressions, sorted by congruence (match/mismatch) and presentation type (analog/linear). Error bars reflect standard error.

Figure 4 shows the response time data, sorted by text-emoji congruence and color-coded by presentation type.

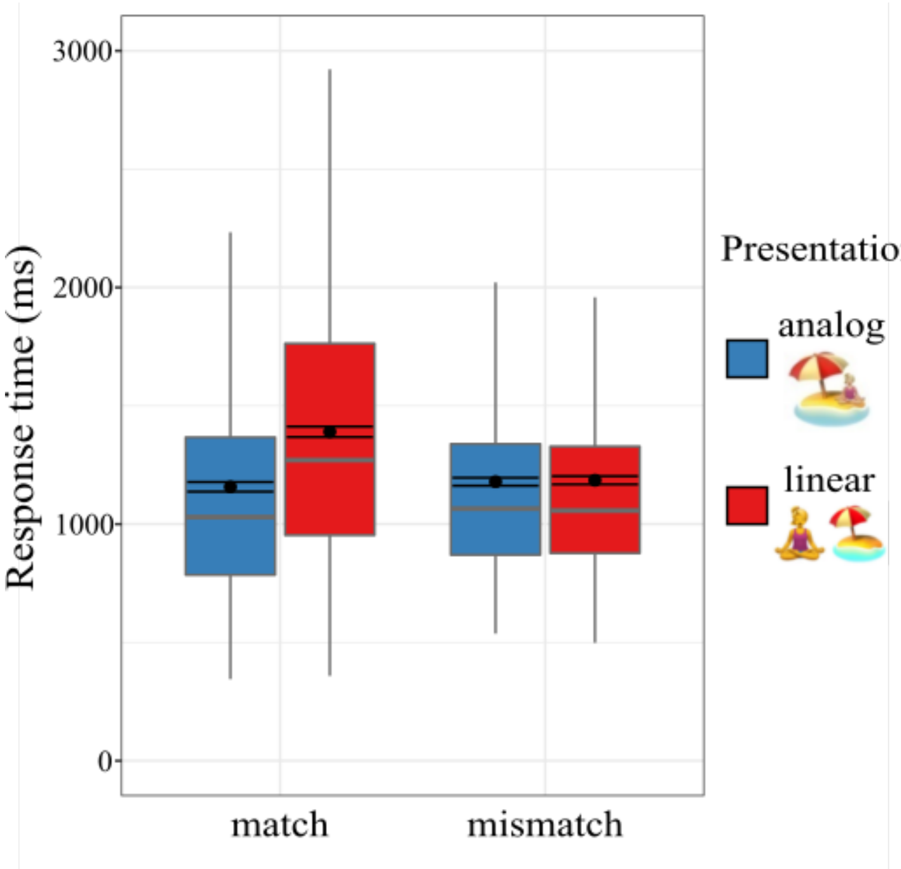


Figure 4. Response times to emoji expressions, sorted by congruence (match/mismatch) and presentation type (analog/linear). Grey horizontal lines depict medians, black circles depict means with standard error bars, vertical whiskers extend to 1.5*IQR.

The results from the aforementioned linear mixed effects model, with Congruence, Presentation, and their interaction as predictors, are presented in Table 1.

	Estimate	**SE**	**t**	***p***
Intercept	1373.9	35.19	39.04	< .001 ***
Congruence[Congruent]	78.71	12.55	6.27	< .001 ***
Presentation[Analog]	-95.71	12.46	-7.68	< .001 ***
Congruence[Congruent]: Presentation[Analog]	-87.18	12.48	-6.99	< .001 ***

Note: Rows correspond to the model estimate, Standard Error, *t*-value, and *p*-value for the factor level specified in the leftmost column. Contrasts were set with sum coding.

Table 1. Summary of linear mixed effects regression modeling response time based on text-emoji congruence and presentation

There was an interaction between congruence and presentation type (t = -7.00, *p* < .001). As determined by post-hoc estimated marginal means via *emmeans* (Lenth, 2018), response times did not differ significantly between the two types of presentation for images that mismatched their text (p = 0.63); we did, however, observe a difference for those that matched (p < .001). For stimuli featuring a match between the phrase and the emoji depiction, the linear presentation yielded significantly slower response times than the analog pictures.

Additional analyses were conducted, including the time participants spent on the first untimed screen (henceforth Reading Time) as a covariate (with outliers removed). Whether raw Reading Time or Reading Time z-scored by participant is used, there were no significant interactions between Reading Time and any of the variables of interest; with Reading Time included as a covariate, there is still a significant interaction between Congruence and Presentation in the same direction as when this covariate is omitted.

Discussion

In this experiment we questioned the affordances of emoji by contrasting them in linear strings against analog pictures. Presenting emoji sequences in a linear order as if they were textual words significantly decreased the response accuracy and increased the response time compared to an analog combination of the same emoji. These results suggest that such a forced linear sequence is not as natural for participants, hinting at the combinatorial parameters of emoji and their structural affordances.

The results of this experiment suggest limitations on the way that compositionality manifests in the graphic modality: We observed that a linear presentation of multi-emoji phrases incurs a processing cost compared to a nonlinear, analog presentation. For example, as in the example in Figure 1, for a sentence like “the girl did yoga at the beach,” the linear presentation (girl doing yoga emoji next to beach emoji) yielded slower response times and lower accuracy rates than an analog presentation (girl doing yoga emoji positioned as if she were *on* the beach emoji). This lends support to the idea that graphics are not optimized for the linear presentation of compositional meaning utilized in text. In this example, it is apparent that forcing graphics into a linear order like text hinders their interpretability compared to the analog presentation of holistic pictures. These results are consistent with the idea that the ways different modalities may afford compositionality is modulated by the ways in which they present information, even though big-picture principles of semantic compositionality and memory themselves do not change across different domains (e.g., Federmeier & Kutas, 2001; Ganis et al., 1996; Hamm et al., 2002; Willems et al., 2008). Further levels of nuance are likely warranted here, including the different possible semantic relations between emoji in expressions like these; further research should explore the processing of different relation types systematically to gain further nuance into the meaning-making processes for emoji sequences.

That this effect was found specifically in the match condition and not in the mismatch condition further supports the notion that combining emoji in a linear sequence is less afforded by the graphic modality. When the preceding text does not match the emoji, participants are tasked merely with identifying component parts and recognizing that the emoji had no alignment in meaning with the sentence. Assessing congruency would thus not require combining the elements or determining their relation. However, in the case of matching text-emoji relationships, recognition of meaning should be equally accessible for both the linear and analog presentations, as they ultimately contained the exact same component parts. Yet, accuracy was higher and response times were faster for analog presentations than linear presentations. This suggests that the graphic modality better affords combinatorial meaning-making in an analog representation than in a linear string.

Given that a linearly presented emoji sequence may leave relations between pictures inexplicit, it demands various inferences to resolve the relations between the units. The example given earlier of 🧘‍♀️⛱️ may be interpreted as positional ("the girl did yoga at the beach"), sequential ("the girl did yoga and then went to the beach"), or merely a list ("there was a girl, there was a beach"). Depending on this inference, the likelihood of a "match" response might thus vary for the linear presentations. This range of inferencing is precisely one reason why emoji indeed lack the affordances of a linear presentation. Spoken language – and the analog depictions from the present experiment – are capable of being more explicit about these meaning relations, motivating the idea that these expressions can express a proposition like "the girl did yoga at the beach" more unambiguously than can linear emoji sequences.

Though our findings suggest that linear sequencing indeed is less optimal for pictorial representations, other experimental approaches would address some of the limitations of the present study and add further clarity and nuance to this finding. A free-response production task on these (or similar) stimuli in both their linear and analog presentations would elucidate the meaning-making processes involved in each by exploring participant-generated "translations" from the image depictions themselves rather than matching to a provided phrase. Based on the findings here, we would hypothesize that participants would generate greater consistency of interpretations for analog emoji combinations than linear ones. A systematic manipulation of different types of meaning relations in a linear sequence could also clarify whether certain meaning relations or emoji types may lend themselves better to the linear sequencing format than others. Lastly, it should be noted that this experiment was conducted in English only. It is possible that systematic cultural, linguistic, or writing system differences may affect results in a study like this. A similar exploration conducted in another language (such as, given the findings of Herring & Ge [2020], Chinese) would elucidate the specificity/generality of the analog preference observed in the current experiment.

This finding may shine some light on approaches like that described in Vandeghinste et al. (2017) that aim to translate text into pictographs. Such an effort converts text input into a pictographic output, offering purported communicative advantages to illiterate individuals. Indeed, evidence suggests pictorial sequences attempting to translate text appear to be parasitic on the word order of a reader's speech (Nakamura et al., 2006), not a form of universal communication. Given the less natural disposition of forced-linear presentations of pictographs found in the present study, such translations may offer no communicative advantage over a traditional analog pictorial representation of information. Moreover, our findings suggest that the technological limitations imposed on emoji, such as their forced linearity and fixed sizing, may be suboptimal for the affordances of their modality, despite the clear usefulness of emoji in online communication.

It is worth noting that our analysis of compositionality here arguably does not even approach the issue of emoji grammar. We examined no aspects of parts-of-speech, phrase structures, or other canonical aspects of syntactic structure which have been shown to be characteristically absent in the production of emoji sequencing (Cohn et al., 2019, cf. Herring & Ge, 2020). Rather, our analysis here examines the even simpler issue of the combinatorial semantics of multi-emoji messages. Even for a simple pairing of emoji, we find semantic compositionality to be strained compared to a more naturalistic pictorial representation that renders the presentation into a more conceptually dense single unit.

In addition, though we observed here a disadvantage for emoji conveying information in a linear way, it does not discount the general potential for the graphic modality to convey information across a complex sequence displaying grammatical properties like categorical roles ("parts of speech") or hierarchical recursion. Visual sequences have indeed shown grammatical properties at the narrative level, such as in the sequencing of visual narratives like comics, where each sequential unit is a compositionally whole visual scene (Cohn, 2013a). Such sequences allow for recursive embedding of units with categorical roles, giving way to sequences with center-embedded clauses, structural ambiguities, long distance dependencies, and anaphoric relationships (Cohn, 2013a; Coopmans & Cohn, 2022). Emoji then may simply be limited at the level of information structure in what they convey per unit within a linear sequence. Visual information at a "word" level may not be optimized for linear sequencing, while scene-level pictures that convey more information per unit could better afford the complexity of a full grammatical system.

On this view, modalities have affordances for different levels of information structure (see Cohn, 2013b), and drawing direct equivalences between how modalities manifest their abstract structuring may inappropriately funnel a modality into doing something it is not naturally optimized for. In other words, the graphic modality does have the potential for similar properties of linguistic structure as the verbal modality (i.e., lexical entrenchment, compositionality, hierarchy, etc.), but such structures may manifest in ways that differ on the basis of the affordances of the modalities themselves (Cohn, 2013a). Forcing the graphic modality to operate in a less natural way, like linear strings of emoji, thus constrains the communicative capacity of the modality. More generally, the evidence presented here seems to support Cohn and Schilperoord's (2023) response to Morin (2023): Coercing graphic modality content into the structures of the spoken modality is cognitively disadvantageous, and general, robust, self-sufficient ideographies built in this way are highly unlikely to ever develop.

A byproduct of such a constraint may indeed be that emoji act more like co-speech gestures or emblems than fully combinatorial language itself (e.g., Cohn et al., 2019; Grosz et al., 2021; McCulloch & Gawne, 2018), since their use in context primarily optimizes simple, single-emoji expressions or basic morphological strategies like repetition.

Interestingly, since this experiment was run, a project called Emoji Kitchen (2023) has been released and grown in popularity. This tool allows users to create mashups from two existing Android emoji – not every pair will work, but it sponsors a variety of mashup types. For example, the classic face with tears of joy emoji 😂 can be merged with a number of emoji to create a skull with tears of joy, koala with tears of joy, lemon with tears of joy, spaghetti with tears of joy, saxophone with tears of joy, the recycling logo with tears of joy, and hundreds of others.

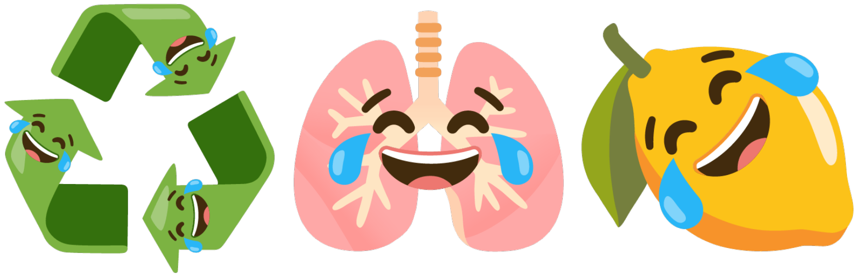


Figure 5. Recycling logo with tears of joy, lungs with tears of joy, lemon with tears of joy (Emoji Kitchen, accessed via https://emojikitchen.dev/)

Though only a subset of emoji appear in this workspace and only a subset of those possible combinations exist, this represents a step in the direction of picture-like, naturalistic multi-emoji expressions. At the time of writing, none of the "activity" emoji (e.g., a woman doing yoga) are part of the Emoji Kitchen, no iOS emoji version yet exists, and most of the creations (e.g., recycling logo with tears of joy) are likely more useful for a chuckle than for expressing full-sentence meaning or allow for more naturalistic visual sequencing. iOS17 has also introduced a way to layer whole emoji as stickers to compositionally create single images (where a woman doing yoga emoji could be placed on a beach emoji), but without the ability to blend them like Emoji Kitchen. Time will tell whether these tools evolve into a more accessible and stable part of the emoji ecosystem.

Thus, while investigating the linguistic structure of emoji can offer insights into their communicative potential, it remains important to consider the affordances of the modalities themselves being compared. In the case of emoji and other ideographic systems, this sensitivity to the nature of graphics extends to whether technology constrains or facilitates those affordances.

Conflict of Interest

Neil Cohn has advised in the creation of new emoji and in commercial use of emoji.

References

Barach, E., Feldman, L. B., & Sheridan, H. (2021). Are emojis processed like words?: Eye movements reveal the time course of semantic processing for emojified text. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/s13423-020-01864-y

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., …

Grothendieck, G. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beyersmann, E., Wegener, S., & Kemp, N. (2023). That's good news ☹: Semantic congruency effects in emoji processing. *Journal of Media Psychology: Theories, Methods, and Applications*, *35*, 17–27. https://doi.org/10.1027/1864-1105/a000342

Brehm, L., & Alday, P. M. (2022). Contrast coding choices in a decade of mixed models. *Journal of Memory and Language*, *125*(January), 104334. https://doi.org/10.1016/j.jml.2022.104334

Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.

Cohn, N. (2013a). *The visual language of comics: Introduction to the structure and cognition of sequential images*. Bloomsbury.

Cohn, N. (2013b). Visual narrative structure. Cognitive science, 37(3), 413-452.

Cohn, N. (2016). A multimodal parallel architecture : A cognitive framework for multimodal interactions. *Cognition*, *146*, 304–323. https://doi.org/10.1016/j.cognition.2015.10.007

Cohn, N. (2018). Combinatorial morphology in visual languages. In Booij, Geert (Ed.) *The Construction of Words: Advances in Construction Morphology*, 175–199. London: Springer.

Cohn, N., Engelen, J., & Schilperoord, J. (2019). The grammar of emoji? Constraints on communicative pictorial sequencing. *Cognitive Research: Principles and Implications*, *4*(1). https://doi.org/10.1186/s41235-019-0177-0

Cohn, N., & Foulsham, T. (2022). Meaning above (and in) the head: Combinatorial visual morphology from comics and emoji. *Memory and Cognition*. https://doi.org/10.3758/s13421-022-01294-2

Cohn, N., Murthy, B., & Foulsham, T. (2016). Meaning above the head: Combinatorial constraints on the visual vocabulary of comics. *Journal of Cognitive Psychology*, *28*(5), 559–574. https://doi.org/10.1080/20445911.2016.1179314

Cohn, N., & Schilperoord, J. (2022). Remarks on multimodality: Grammatical interactions in the parallel architecture. *Frontiers in Artificial Intelligence*, *4*(January), 1–21. https://doi.org/10.3389/frai.2021.778060

Cohn, N., & Schilperoord, J. (2023). Visual languages and the problems with ideographies: A commentary on Morin. *Behavioral and Brain Sciences*, 26–28.

Coopmans, C. W., & Cohn, N. (2022). An electrophysiological investigation of co-referential processes in visual narrative comprehension. *Neuropsychologia*, *172*. https://doi.org/10.1016/j.neuropsychologia.2022.108253

Culicover, P. W., & Jackendoff, R. (2005). *Simpler syntax*. Oxford University Press.

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

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12.

Emoji Kitchen. (n.d.). Retrieved from 2023 website: https://emojikitchen.dev/

Federmeier, K. D., & Kutas, M. (2001). Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *27*(1), 202–224. https://doi.org/10.1037//027

Feldman, L. B. (2023). Emoji use validates the potential for meaning standardization among ideographic symbols. *Behavioral and Brain Sciences*, 29–30.

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for "common sense": An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience*, *8*(2), 89–106.

Gawne, L., & McCulloch, G. (2019). Emoji as digital gestures. *Language@ Internet*, *17*(2).

Ge, J., & Herring, S. C. (2018). Communicative functions of emoji sequences on Sina Weibo. *First Monday*, *23*(11). https://doi.org/10.5210/fm.v23i11.9413

Grosz, P., Kaiser, E., & Pierini, F. (2021). Discourse anaphoricity and first-person indexicality in emoji resolution. *Proceedings of Sinn Und Bedeutung*, *25*, 340-357.

Hamm, J. P., Johnson, B. W., & Kirk, I. J. (2002). Comparison of the N300 and N400 ERPs to picture stimuli in congruent and incongruent contexts. *Clinical Neurophysiology*, *113*(8), 1339–1350.

Herring, S., & Ge, J. (2020). Do emoji sequences have a preferred word order? In *Proceedings of the 3rd International Workshop on Emoji Understanding and Applications in Social Media (Emoji2020).* DOI: 10.36190/2020.05

Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences*, *23*(8), 639–652. https://doi.org/10.1016/j.tics.2019.05.006

Holtgraves, T., & Robinson, C. (2020). Emoji can facilitate recognition of conveyed indirect meaning. *PLoS ONE*, *15*(4), 1–13. https://doi.org/10.1371/journal.pone.0232361

Homann, L. A., Roberts, B. R. T., Ahmed, S., & Fernandes, M. A. (2022). Are emojis processed visuo-spatially or verbally? Evidence for dual codes. *Visual Cognition*, *30*(4), 267–279. https://doi.org/10.1080/13506285.2022.2050871

Kaye, L. K., Darker, G. M., Rodriguez-Cuadrado, S., Wall, H. J., & Malone, S. A. (2022). The Emoji Spatial Stroop Task: Exploring the impact of vertical positioning of emoji on emotional processing. *Computers in Human Behavior*, *132*(October 2021), 107267. https://doi.org/10.1016/j.chb.2022.107267

Kaye, L. K., Rodriguez-Cuadrado, S., Malone, S. A., Wall, H. J., Gaunt, E., Mulvey, A. L., & Graham, C. (2021). How emotional are emoji?: Exploring the effect of emotional valence on the processing of emoji stimuli. *Computers in Human Behavior*, *116*(November 2020), 106648. https://doi.org/10.1016/j.chb.2020.106648

Langacker, R. W. (1987). *Foundations of cognitive grammar: Descriptive application. Volume 2* (Vol. 2). Stanford University Press.

Lenth, R. (2018). Emmeans: Estimated marginal means, aka least-squares means. *R Package Version*, *1*(1).

McCulloch, G., & Gawne, L. (2019). Emoji grammar as beat gestures. *Language@Internet*, *17*(2).

Meloni, L., Hitmeangsong, P., Appelhaus, B., Walthert, E., & Reale, C. (2022). Beyond emojis : an insight into the IKON language. In *Proceedings of the Fifth International Workshop on Emoji Understanding and Applications in Social Media*, 11–20.

Morin, O. (2023). The puzzle of ideography. *Behavioral and Brain Sciences*. https://doi.org/10.1017/S0140525X22002801

Nakamura, K., Iwabuchi, M., & Alm, N. (2006). A cross-cultural study on the interpretation of picture-based sentences. *International Journal of Computer Processing of Languages*, *19*(04), 239–248. https://doi.org/10.1142/s0219427906001529

Norré, M., Vandeghinste, V., Bouillon, P., & François, T. (2021). Extending a text-to-pictograph system to French and to Arasaac. *International Conference Recent Advances in Natural Language Processing, RANLP*, 1050–1059. https://doi.org/10.26615/978-954-452-072-4_118

Paggio, P., & Tse, A. P. P. (2022). Are emoji processed like words? An eye-tracking study. *Cognitive Science*, *46*(2). https://doi.org/10.1111/cogs.13099

Pasternak, R., & Tieu, L. (n.d.). *Co-linguistic content projection: From gestures to sound effects and emoji*. Retrieved from https://ling.auf.net/lingbuzz/005082

Scheffler, T., Brandt, L., Fuente, M. de la, & Nenchev, I. (2022). The processing of emoji-word substitutions: A self-paced-reading study. *Computers in Human Behavior*, *127*(August 2021), 107076. https://doi.org/10.1016/j.chb.2021.107076

Sevens, L., Vandeghinste, V., Schuurman, I., & Eynde, F. Van. (2017). Simplified text-to-pictograph translation for people with intellectual disabilities. *International Conference on Applications of Natural Language to Information Systems*, 185–196.

Vandeghinste, V., Sevens, I. S. L., & Van Eynde, F. (2017). Translating text into pictographs. *Natural Language Engineering*, *23*(2), 217–244. https://doi.org/10.1017/S135132491500039X

Võ, M. L. H. (2021). The meaning and structure of scenes. *Vision Research*, *181*(August 2019), 10–20. https://doi.org/10.1016/j.visres.2020.11.003

Weissman, B., Engelen, J., Baas, E., & Cohn, N. (2023). The lexicon of emoji? Conventionality modulates processing of emoji. *Cognitive Science*, *47*(4). https://doi.org/10.1111/cogs.13275

Weissman, B., & Tanner, D. (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

Wicke, P., & Cunha, J. M. (2020). An approach for text-to-emoji translation an approach for text-to-emoji translation. *ICCC2020: International Conference on Computational Creativity*, (September).

Willems, R. M., Özyürek, A., & Hagoort, P. (2008). Seeing and hearing meaning: ERP and fMRI evidence of word versus picture integration into a sentence context. *Journal of Cognitive Neuroscience*, *20*(7), 1235–1249.

Wright, K. E., Zimmer, B., Carson, C. E., Hughes, B., McLean, J., & Zhang, L. (2023). Among the new words. *American Speech*, *98*(3), 296–317. https://doi.org/10.1215/00031283-2077633

Zipf, G. K. (1935). *The psychobiology of language*. Houghton-Mifflin.

Supplementary Materials

https://osf.io/tq5us/?view_only=0d806f7163e24f1e8084ab4b3a4177cd

This link contains data from the experiment, R code used in analysis, the stimulus set, and the Emoji Language Fluency questionnaire.

Biographical Notes

Benjamin Weissman [weissb2@rpi.edu] is a Lecturer in the Department of Cognitive Science at Rensselaer Polytechnic Institute. His research explores the real-time processing of meaning across a range of linguistic and communicative phenomena, including, especially, emoji.

Jan Engelen [j.a.a.engelen@tilburguniversity.edu] is Assistant Professor in the Department of Communication and Cognition at Tilburg University. His research interests include embodied cognition and language comprehension at the sentence and text level.

Lena Thamsen is a former master's student at Tilburg University, Netherlands. She graduated in 2019 in the study program Business Communication and Digital Media. In her master thesis she evaluated the impact of linearity on emoji sequences.

Neil Cohn [neilcohn@visuallanguagelab.com] is Associate Professor of Communication and Cognition at Tilburg University. He studies the linguistic structure and (neuro)cognition of graphic and multimodal communication, particularly emoji and the visual languages used in comics.