How Many Monkeys Do You See?
Eye-tracking of Visual Co-Reference in Emoji Sequences

Lenneke Doris Lichtenberg
Snr: 2024033

Research Master's thesis Linguistics and Communication Sciences

School of Humanities and Digital Sciences
Tilburg University, Tilburg

Supervisor: dr. Neil Cohn
Second reader: dr. Rein Cozijn

July 2024

Short title: Co-reference in emoji sequences

Acknowledgements

I would like to thank my main academic supervisor and mentor, Dr. Neil Cohn, for his guidance, insights, and support throughout my research master's (thesis). I highly appreciate his trust in granting me the complete freedom to design and conduct research that align with my own personal interests, as well as his support of my decisions. Furthermore, I would like to extend my thanks to my second supervisor, Dr. Rein Cozijn. His flexibility, willingness to help and prompt responses enabled me to quickly learn how to conduct eye-tracking research, a skill that will undoubtedly be valuable in my future research.

Abstract

In the three emoji sequence of 🙈🙉🙊, people differ in their interpretation of whether there are three different monkeys or one monkey changing postures. Why is this the case? I investigated this issue of continuity in visual sequencing by examining the factors influencing people to (not) construe co-reference between emoji in sequences. Participants judged the number distinct referential entities in sequences of three emoji that either suggested a specific order or not, while their eye movements were tracked. Sequences varied in whether emoji belonged to the same superordinate semantic category and/or shared colours. Results showed that participants were more likely to construe co-reference, made fewer switches between emoji, and spent less time looking at emoji that were part of sequences suggesting a sequential order compared to unordered sequences. Participants were also more inclined to construct continuity when emoji in sequences had similar superordinate semantic categories or colours. Moreover, emoji sequences with more ambiguous interpretations induced more switching eye movements between the emoji. Participants' experience with emoji sequences modulated these findings, emphasizing the importance of expertise in co-referential sequential processing. These findings suggest that the understanding of sequential images is not universally transparent, and that construal of continuity across images is constrained by the interplay between visual features and individual experience.

Keywords: emoji, visual language, visual sequences, co-reference, continuity, eye-tracking

Index

1. Introduction

People use more and more emojis in interactive digital communication, which has become increasingly important all over the world (Dürscheid & Haralambous, 2021). As a result of this rising popularity, more attention has been paid to emoji within academia. For example, prior work showed that people adopted several characteristics from verbal language usage when using emoji (Ge & Herring, 2018; Pohl, Domin, & Rohs, 2017). Moreover, others proposed that emoji themselves have developed into a separate language (Scall, 2015; Monti et al., 2016), or that, due to their cross-cultural emergence, they even constitute a universal language (Ai et al., 2017; Evans, 2017).

Yet, not all research on emoji supports this idea of universality (Ge & Herring, 2018; Lichtenberg, Hacimusaoğlu, Klomberg, Schilperoord, & Cohn, In prep). A well-known example that challenges the universal nature of emoji is known as "the monkey emoji debate" (Wilde, 2019). This monkey emoji debate was caused by the question wheter the "three-wise-monkey" emoji (see Figure 1) represent three distinct monkeys (interpretation 1) or just one monkey performing three different poses (interpretation 2). Given that people maintain both interpretations roughly equally, this sequence is ambigous and the perception of emoji sequences is thus not universally transparent (Wilde, 2019). Here, I aim to further explore which factors motivate people to form different interpretations about the number of distinct entities in emoji sequences, like the "three-wise-monkey" emoji, and how this is reflected in their eye movements.

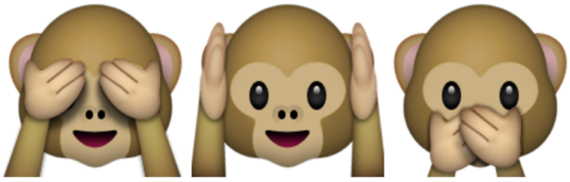

Figure 1. Three wise monkey emoji.

2. Theoretical framework

2.1. The continuity constraint

The inferences people make about the relationship between the monkeys in Figure 1 underlie their perception of whether they see the sequence as one or three monkey(s). When people view an image-unit belonging to a visual sequence, they predict which upcoming semantic information could also be present in the subsequent units (Cohn, 2019). One such basic prediction includes that the same referential entities can be found across the next image-units. This prediction arises from a *continuity constraint* where the same character, rather than a different character, is depicted repeatedly in images within a sequence (Bornens, 1990; Cohn, 2020b; Klomberg et al., 2023).

In other words, the characters are recognized as co-referential, despite physically different lines appearing across images. In case of the monkey emoji, maintaining the continuity constraint would render the interpretation of one monkey that changes postures across three emoji (i.e. all emoji co-refer to the same monkey). On the other hand, a lack of continuity would lead to the interpretation of three separate emoji. Co-reference is thus also required to resolve common entities in visual sequences (Klomberg et al., 2023), despite often being considered a specific property of spoken and signed language (Chomsky, 1980; Gordon & Hendrick, 1997; Graesser, Millis, & Zwaan, 1997; Sanders & Gernsbacher, 2004).

2.2. Similarities aid visual co-reference

Although having different properties, the need to connect distinct units to be one and the same entities is shared by the comprehension of the verbal and visual modality (Klomberg et al., 2023). This is, for example, illustrated by research showing that constraints operating on referential units in language (i.e. anaphora) also apply to visual referential units (i.e. comic panels) (Cohn, van Middelaar, Foulsham, & Schilperoord, 2024). Following linguistic research (Gibson, 2000), it was found that distance constrained co-referential processing in visual sequences, as longer distances between referential units (i.e. comic panels) resulted in slower viewing times and neurocognitive costs.

Consequently, linguistic theories about co-reference have been instrumental in understanding how people construe visual co-reference, such as in Klomberg et al. (2023), who formalized constraints underlying visual co-referencing by utilizing Jackendoff's (1983, 2010) theory of Conceptual Semantics. This theory posited that similarities in the visual representation of sequential units motivate conceptual continuity between entities across these units (e.g. separate emoji), through which visual co-reference arises.

These similarities in the visual representation, for example, include, entities sharing visual features such as colours, contours or poses. According to Klomberg et al. (2023), such similarities aid continuity by prompting the comprehender to connect them across units as common entities. A behavioral experiment from Lichtenberg et al. (In prep) in which participants were asked to judge the number of distinct entities represented in sequences of three emoji supported this idea. Their results indicated that people were more likely to construct continuity if the emoji in sequences shared similar colours and/or belonged to the same superordinate semantic category with similar contours (e.g. being all monkeys).

Even though the “three-wise-monkey” emoji exhibit similar colours and contours, they also represent simple monkey visualizations without individual-specific features. This then contributes to the ambiguity of two interpretations: “the exact same monkey” vs. “three different monkeys”. In a first experiment exploring these types of sequential ambiguities, Lichtenberg et al. (In prep) found that people took more time to judge the number of distinct referential entities of emoji sequences with more ambiguous interpretations, indicating that such sequences may require more cognitive resources to process. They also found that more ambiguous sequences were judged to represent more distinct referential entities.

2.3. Similarities also highlight differences

However, Klomberg et al. (2023) posit that similarities across visual units cannot always convince comprehenders to construct continuity between them, as illustrated by the ambiguous interpretation of the “three-wise-monkey” emoji. This might be explained by the notion that it is also easier to indicate differences between units that share similarities (verbal example: motel vs. hotel, visual example: blue car vs. red car) than between units that differ in their fundamental features (verbal example: motel vs. rocket, visual example: car vs. table) (Gentner & Markman, 1994; Markman & Gentner, 1996).

In their work about similarity judgements, Gentner and Markman (1994) and Markman and Gentner (1996) refer to this as alignable differences vs. non-alignable differences. Alignable differences are highlighted by features that can be compared directly because they share a commonality (e.g. the colour difference between two cars or the different poses of the monkey emoji). In contrast, non-alignable differences are based on fundamentally different features highlighting their distinct nature (e.g. the difference between a car and a table). As findings suggest that people can easier identify alignable than non-alignable differences between units (verbal: Gentner & Markman, 1994; visual: Markman & Gentner, 1996), it is posited that alignable differences play a more focal role when people make similarity judgements.

2.4. Ordered vs. unordered visual sequences

Continuity can also contribute to sequential properties of visual sequences such as the nature of the sequence (Cohn, 2020a; Lichtenberg et al., In prep). Visual sequences either have an *ordered* or *unordered* nature. In ordered sequences, the content of separate image-units motivates them be comprehended as being in a specific order. For instance, the sequence in Figure 2a displays the Christmas journey of a pine tree. People are required to make a connection (and thus construe co-reference) between the separate emoji, before they can infer this Christmas journey of a pine tree. The idea that image-units (e.g. the different emoji) are sequenced in a specific order is often guided by recognition of continuity between entities across these image-units. The “three-wise-monkey” emoji in Figure 1 are also an example of an ordered sequence, as the monkey emoji mimic the poses of monkeys in a Japanese pictorial maxim (Wilde, 2019). In contrast, visual lists in which separate image-units are not required to be depicted in a specific order (Cohn, 2020a). This is exemplified by the heart emoji in Figure 2b, as they do not suggest a sequential order. The distinction between ordered and unordered emoji sequences was supported by Lichtenberg et al. (In prep), who found that changing the positions of the emoji only violated ordered sequences but not unordered ones.

Figure 2. Examples of ordered (a) and unordered (b) emoji sequences.

Prior work on the use of emoji in sequences indicated that people more frequently use unordered or repeated emoji sequences compared to ordered emoji sequences (Cohn et al., 2019; Gawne & McCulloch, 2019; McCulloch & Gawne, 2018; Tatman, 2018). Despite being a less prevalent type of emoji sequencing, a behavioral experiment by Lichtenberg et al. (In prep) indicated that ordered sequences were easier to judge in terms of represented referential entities compared to unordered sequences, as indicated by faster response times. They also showed that people were more likely to construe co-reference between emoji if the sequence suggested a sequential order compared to unordered sequences.

2.5. The activity constraint

In addition to continuity, people also make inferences about changes that occur between entities across two image-units. As posited by an *activity constraint*, certain alterations in visual depiction can cue shifts in time, viewpoint, or causation. These shifts may induce changes occurring across entities in two image-units, regardless of continuity (Bornens, 1990; Cohn, 2020a; Cohn, 2020b; Klomberg et al., 2023). For instance, the changing positions of the monkey's hands in Figure 1, suggest that, when continuity is maintained, a singular monkey moved its hands across the image-units. Sequential image comprehension thus requires both the continuity and activity constraint (Cohn, 2020a).

Without referential continuity, each image-unit depicts a different entity in the sequence. On the other hand, continuity without activity yields the interpretation that the sequence features the same entity in unrelated scenes (i.e. the monkeys in Figure 1 depict the same monkey, but in unconnected scenes). The lack of both continuity and activity suggests that separate characters in unrelated scenes are depicted across image-units. As both constraints operate across different visual sequence types (i.e. panels in comics or sequenced emoji), they are considered to be fundamental aspects required for sequential image comprehension (Bornens, 1990; Cohn, 2020b; Klomberg et al., 2023).

2.6. The importance of exposure

While it may seem intuitive to maintain continuity in visual sequences, extensive research has shown that various people do not construe this sequential nature (Cohn, 2019). The continuity constraint is often not construed by people from rural areas who lack exposure to Western literacy and culture (Bishop, 1977; Fussell & Haaland, 1978; Byram & Garfoth, 1980; Cook, 1980; San Roque et al., 2012; Gawne, 2016). In these cases, comprehenders view each image in a visual sequence as an independent scene with different characters, instead of construing referential continuity. Such findings dispute the universal transparency of sequential images and suggest that exposure to visual narratives is necessary for people to comprehend them as sequences (Cohn, 2020a). This parallels the requirement of exposure to a linguistic system to achieve fluency in a spoken or signed language.

Developmental research also demonstrates that exposure plays a fundamental role in learning to connect referential entities across images, enabling full comprehension of visual sequences (Bornens, 1990; Trabasso & Nickels, 1992). Differences in this developmental trajectory can arise because not everyone has the same access to visual narratives, due to factors such as socio-economic status or cross-cultural proficiency differences. Bornens (1990), for example, reported a delayed developmental learning trajectory for less "culturally privileged" children, who are possibly less exposed to visual narratives. Moreover, Nakazawa and Shwalb (2012) found that Japanese college students had more experience with visual narratives than American college students and attributed this disparity to the ubiquitous presence of Manga in Japanese culture versus the niche audience of comic readership in the United States. Experimental research also supports that people's overall experience with visual narratives affects visual narrative processing, as it was found to modulate both behavioral (Cohn, 2020b) and brain responses (Coderre & Cohn, 2023).

Besides effects of general fluency for visual (narrative) sequences, prior work also shows that a specific type of visual fluencies can modulate comprehension of specific visual sequences (Cohn & Kutas, 2017; Lichtenberg et al., In prep). Cohn and Kutas (2017), for example, found that readership of Japanese manga modulated the comprehension of specific narrative structures that appear prevalently in manga. Moreover, Lichtenberg et al. (In prep) found that experience with emoji specifically affected people's judgements and response times when they had to judge the number of distinct referential entities in emoji sequences. Here, people with more emoji experience responded faster and were more likely to construe co-reference between the emoij in sequenes. However, they did not find this effect for people's general proficiency with visual narratives, suggesting that a specific type of visual fluency (e.g. emoji fluency) is required to construct continuity across a specific type of visual sequences (e.g. emoji).

2.7. Eye-tracking as a method to investigate visual co-reference in emoji sequences

A frequently used method to assess higher order comprehension processes involved in sequential image comprehension is eye-tracking (Nakazawa, 2002; Foulsham, Wybrow, & Cohn, 2016; Laubrock, Hohenstein, & Kümmerer, 2018; Tseng, Laubrock, & Pflaeging, 2018; Hutson, Magliano, Loschky, 2018). Eye-tracking allows for directly monitoring participants' attentional focus throughout an (experimental) task without interruption (Karabanov, Bosch, & König, 2007). For that reason, it enables a more detailed registration of cognitive processes that, for example, underly language and image processing, compared to more conventional psycholinguistic measures such as reading time.

Most eye-tracking studies focusing on sequential image comprehension included visual stimuli in the form of comic strips or graphic novels (Nakazawa, 2002; Foulsham, Wybrow, & Cohn, 2016; Laubrock, Hohenstein, & Kümmerer, 2018; Tseng, Laubrock, & Pflaeging, 2018; Hutson, Magliano, Loschky, 2018). In only one of these studies, the order of the visual units in the sequence was manipulated (Foulsham et al., 2016). In this study, participants saw comic strips of which the comic panels were either presented in the original or randomized order. Their results indicated that randomizing such ordered sequences affects people’s eye movements and comprehension, as people needed fewer fixations to understand comic panels in the original order more quickly compared to the randomized condition. They also found that comic panels presented in a randomized order required more regressive eye movements and attention.

Moreover, none of the above mentioned studies used eye-tracking to asses visual sequences of emoji. Eye-tracking studies on emoji processing mainly focused on word substituion by emoji and/or the addition of emoji to sentences (Scheffler, Brandt, de la Fuente, & Nenchev, 2022; Paggio & Tse, 2022; Robus, Hand, Filik, & Pitchford, 2020; Barach, 2021). Properties of the emoji, such as ambiguity and congruency with the sentence context, affected people’s eye movements (Paggio & Tse, 2022; Barach, 2021). Paggio and Tse (2022) showed that emoji with ambiguous interpretations elicited longer first fixation durations, longer total viewing durations and longer regression durations. Furthermore, sentences followed by emoji that were incongruent with the sentence context evoked longer fixations compared to synonymous emoji (Barach, 2021). Brain responses to incongruent emoji substituting a word in a sentence also showed sustained costs of semantic processing (N400) compared to congruent emoji (Weissman, 2019). In contrast to emoji combined with sentences, less attention has been payed to people’s eye movements while viewing emoji sequences.

As previous work using eye-tracking to investigate visual sequencing and emoji processing did not focus on visual co-reference, eye-tracking studies that assessed verbal co-reference can be informative. Psycholinguistic studies found that discontinuity and ambiguity between anaphors (i.e. words/phrases that refer back to earlier words/phrases) and antecedents (i.e. words/phrases being referred to) were reflected in people’s eye movements (Rayner, Chace, Slattery, & Ashby, 2006; Cook, 2005, cited in Rayner, 2006; Spivey-Knowlton & Tanenhaus, 2015). Anaphors inconsistent with their antecedents (e.g. “carrot sticks” used to refer to “celery sticks”) yielded longer reading times, longer fixations on the anaphor and more regressions to the antecedent (Rayner et al., 2006). Increased reading times and regressions were found when there was both low (e.g. cello-oboe) and high (e.g. cello-violin) semantic overlap between inconsistent anaphors and antecedents (Cook, 2005, cited in Rayner, 2006). Similarly, ambiguous anaphors induced longer reading times and more regressions compared to unambiguous anaphors (Spivey-Knowlton & Tanenhaus, 2015).

2.8. The present study

Several factors can thus contribute to people (not) construing co-reference between image-units within a visual sequence. Where some of these factors are bound to the (entities in the) sequence itself (i.e. sequence nature and visual information such as colour and categorical membership), can others be attributed to the comprehender of the sequence (e.g. prior exposure and experience with visual narratives). Prior work mainly assessed the sequential properties of emoji via behavioral experiments (Cohn et al., 2019; Lichtenberg et al., In prep), which provide a less detailed registration of cognitive processes compared to eye-tracking measures (Karabanov et al., 2007). Prior eye-tracking research that allowed for assessing these higher order comprehension processes during sequential image processing mostly studied comic strips and graphic novels (Nakazawa, 2002; Foulsham, Wybrow, & Cohn, 2016; Laubrock, Hohenstein, & Kümmerer, 2018; Tseng, Laubrock, & Pflaeging, 2018; Hutson, Magliano, Loschky, 2018). Furthermore, prior work that investigated emoji processing via eye-tracking mainly focused on word substitution by emoji or emoji additions to sentences (Scheffler, Brandt, de la Fuente, & Nenchev, 2022; Paggio & Tse, 2022; Robus, Hand, Filik, & Pitchford, 2020; Barach, 2021). Little research, however, investigated the higher-level comprehension processes involved in the processing of emoji sequences, desipte their global prevalence and prominent social discussion about continuity, like the monkey emoji debate (Wilde, 2019).

For that reason, I aim to assess how factors at play in construing (or not construing) continuity between emoji in sequences affect processing via eye-tracking. Participants viewed ordered and unordered emoji sequences of three emoji which differed in terms of their colour similarity, categorical membership and ambiguity of interpretations. Scrambled versions of each sequence changed the positions of the emoji. The participants judged the number of distinct referential entities represented in each emoji sequence, while their judgements, decision times and eye movements during the viewing of each sequence were measured.

2.9. Hypotheses

Both sequential and entity properties can be influential when people construe co-reference across entities in sequential images. An ordered sequence nature exemplifies a sequential property driven by the continuity constraint (e.g. people need to construe co-reference between the emoji in Figure 2a to infer that the sequence represents the Christmas journey of a pine tree) (Cohn, 2020a). Furthermore, entity properties such as visual similarities (e.g. colour or superordinate semantic category) have been theorized to prompt co-reference (Klomberg et al., 2023). As Lichtenberg et al. (In prep) found that these factors influenced people's judgements about the number of distinct referential entities in emoji sequences and response times, I firstly expect to find similar results and hypothesize that:

H1a: People will judge ordered emoji sequences faster and perceive them to have fewer distinct entities than unordered emoji sequences.
H2a: Greater similarity (i.e. through colour/categorical membership) between emoji within sequences will lead to faster decision times and judgments of fewer distinct entities.

Additionally, prior work that randomized the order of comic panels that formed an ordered visual sequence found that randomized sequences induced more regressive eye movements, longer fixations compared to original sequences (Foulsham et al., 2016). Furthermore, linguistic eye-tracking studies found that continuity between anaphora was reflected in people's eye movements through shorter fixations and fewer regressions, and elicited shorter reading times (Rayner et al., 2006; Cook, 2005, cited in Rayner, 2006). Therefore, I secondly predict that the construction of continuity between emoji in sequences will also be reflected in people's eye movements:

H1b: People will make fewer (regressive) switches between emoji and have shorter emoji viewing times while judging ordered emoji sequences compared to unordered emoji sequences.
H2b: Greater similarity (i.e. through colour/categorical membership) between emoji within sequences will elicit fewer (regressive) switches between emoji and shorter emoji viewing times.

Moreover, as only ordered sequences suggest a specific order while unordered sequences allow for more sequential flexibility (Cohn, 2020a), I expect that, in line with Lichtenberg et al. (In prep):

H3ab: Scrambling the emoji sequences by changing the positions of the emoji will affect people's decision times and judgements only for ordered sequences and not unordered sequences, and this will also be reflected in their eye movements.

Furthermore, even though Klomberg et al. (2023) posit that visual similarities aid co-reference between image-units, their theory also accounts for people sometimes not construing continuity, despite these similarities (as illustrated by disagreement about the interpretation of the “three-wise-monkey” emoji). Given that Lichtenberg et al. (In prep) showed that response times and judgements were modulated by people’s agreement about the distinct referential entities represented in each emoji sequence, I expect to replicate this result and predict that:

H4a: People will judge more ambiguous emoji sequences slower and perceive them to represent more distinct referential entities.

In addition, as eye-tracking research on co-reference in ambiguous sentences showed that people needed more reading time and had more regressive eye movements when they were presented with ambiguous anaphors compared to unambiguous anaphors (Spivey-Knowlton & Tanenhaus, 2015), I predict results of a similar nature for co-reference in ambiguous emoji seqences:

H4b: More ambiguous emoji sequences will evoke more (regressive) switches between emoji and shorter emoji viewing times.

Lastly, I consider that participants’ experience with visual sequences will modulate these effects, as prior work showed that people need to be exposed to visual narratives before they can construe co-reference between entities across image-units (Bishop, 1977; Fussell & Haaland, 1978; Byram & Garfoth, 1980; Cook, 1980; San Roque et al., 2012; Gawne, 2016). Given that prior experimental studies indicated that people’s comprehension of (specific) visual sequences can be modulated by both general visual fluency (Cohn, 2020b; Coderre & Cohn, 2023) and specific visual fluencies (Cohn & Kutas, 2017; Lichtenberg et al., In prep), I expect that:

H5ab: People’s overall experience with visual narratives and specific experience with emoji will affect their decision times and judgements, which will also be reflected in their eye movements.

3. Methods

3.1. Participants

52 students from Tilburg University participated in the study in return for course credit. All participants gave their informed consent following the guidelines of the Tilburg School of the Humanities and Digital Sciences Research Ethics and Data Management Committee. Due to an unsuccessful 9-point calibration (>2° error for x and y axis), nine participants were excluded. Moreover, three additional participants were excluded during the analysis phase, as their eye tracking data contained too much noise. The remained study sample included 40 participants (27 woman, 13 men; mean age 21.25, *SD* = 2.56, range: 18-28). All participants had normal or corrected-to-normal (glasses, contacts) vision.

Prior to experimentation, participants were asked to fill out the Visual Language Fluency Index, a questionnaire that assesses participants' visual language expertise through questions about how often they read and drew visual narratives (e.g. comic books, comic strips graphic novels) (Cohn, 2020a). On average, low fluency is indicated by VLFI scores below 8, average around 12 and high around 22. Participants in this experiment had a mean VLFI score of 16.81 (*SD* = 8.46, range: 4.21-34.38), an average-high proficiency.

In addition, participants filled out the Emoji Language Fluency 2 questionnaire (ELF2) that assesses participants' emoji habits via questions about (sequential) emoji use and reception, enjoyment and efficiency (Lichtenberg et al., In prep). ELF2 supplements the original Emoji Language Fluency Questionnaire (ELF) (Weissman, Engelen, Baas, & Cohn, 2023) with more questions about sequential emoji use and allows participants to better specify their emoji proficiency through more questions about the way they insert emoji into their messages. As ELF2 scores range from 1 (= low) to 7 (= high) fluency, participants in this experiment can be considered as fluent (*M* = 5.03, *SD* = .75, range: 2.64-6.31).

3.2. Materials

The stimuli were identical to the stimuli used in Lichtenberg et al. (In prep), where 60 unique emoji sequences were created, of which 30 were ordered (i.e. the emoji in the sequence were bound to a specific order) and 30 were unordered (i.e. the emoji in the sequence were not bound to a specific order). These 60 emoji sequences were paired with 60 additional emoji sequences in which the positions of the emoji were randomly scrambled. This rendered a scenario with four conditions that manipulated the factors of sequence nature (ordered vs. unordered) and scrambling (non-scrambled vs. scrambled) (Figure 3).

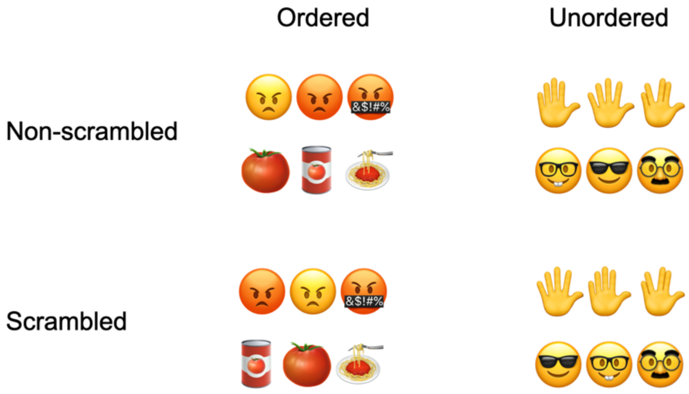

Figure 3. Example sequences across sequence nature and scrambling conditions that differ in terms of categorical and colour similarity scores.

Furthermore, Lichtenberg et al. (In prep) gave each emoji sequence similarity scores for categorical membership (i.e. whether the emoji in the sequence belonged to the same superordinate semantic category or not) and colour. These scores represented the average ratings of pairwise comparisons between all emoji within a sequence (1 = similar, 0.5 = somewhat similar, 0 = not similar). Each sequence was rated by two people and disagreements were solved by discussion. For instance, the "angry face emoji" sequence in Figure 3 was scored a 1 (similar) in terms of categorical similarity (all face emoji) and a 0.17 (not so similar) in terms of colour (given the differences between yellow, red, and black). Moreover, Pearson's item correlations indicated a positive correlation between emoji sequences' categorical and colour similarity ($r(118) = .43$, $p <.001$), indicating that when emoji belonged to the same superordinate semantic category, they were more likely to have similar colours.

3.3. Design

The study included a two (sequence nature: ordered vs. unordered) by two (non-scrambled vs. scrambled) within-subject design, as all participants were exposed to all four emoji sequence types (ordered non-scrambled, unordered non-scrambled, ordered scrambled, unordered non-scrambled). The sequences were distributed into two lists of 60 trials that were counterbalanced so that each participant saw each sequence only once (either non-scrambled or scrambled) and saw 15 trials of each sequence type. Participants were randomly assigned to one of these two lists. Each trial consisted of one sequence of three emoji presented all at once.

3.4. Instrumentation

For each sequence of three emoji, participants had to judge whether they interpreted the sequence as having one, two or three distinct referential entities. The primary dependent eye-movement measures mainly focused on participants' switches (e.g. switching from one area of interest to another area of interest) between the three emoji. These measures included the number of switches between the three emoji, the number of forward switches between the three emoji, the number of backward switches between the three emoji, the number of switches between all possible combinations of two emoji (1-2, 1-3, 2-1, 2-3, 3-1, and 3-2), total viewing time of all three emoji.

Furthermore, participants judgements (either 1, 2 or 3) and the time they took to make these judgements were measured as dependent variables. This was further substantiated with a subsidiary eye movement measure that included the number of switches between emoji and answer options.

3.5. Equipment

Two computers were used for the experiment, one stimulus computer and one eye tracker computer that were connected by USB. The stimuli were presented on a 24-inch monitor. Eye movements of both eyes were sampled horizontally and vertically at a rate of 500 Hz by an EyeLink Portable Duo eye tracker, which had an average accuracy between 0.25-0.5°.

Participants completed the experiment in a soundproof booth. They were positioned in a comfortable chair at a 60cm distance from the eye tracker. The text (font: Calibri, 11-16pt), numbers (font: Calibri: 50pt) and emoji images (covering approximately 400 x 400 pixels) were presented on a white (#FFFFFF) background (1920 x 1080 pixels).

3.6. Procedure

Prior to experimentation, participants were asked to fill out some demographical questions and questions about their experience with comic reading (measured with VLFI) and emoji (measured with ELF2). After this, participants were asked to sit in a comfortable position while they could still use the keyboard. After a successful 9-point calibration (<2° error for x and y axis), participants viewed an instruction screen on the monitor. This included that they would view sequences of three emoji and that they had to decide whether a sequence represented one, two or three referential entities by pressing the corresponding number on the keyboard. To make sure that the experimental task was clear, the experimenter also explained this instruction orally. Furthermore, participants were informed that this was not a reaction time experiment and that they could use all time needed to judge the sequences. This was followed by one practice trial and a short recap of the instruction. After this, the experimenter left the experimental booth.

Each trial began with a screen reading "How many distinct entities do you think are in the next sequence? Please press the corresponding number on the keyboard.". A fixation point in the form of a red dot, which position corresponded to the position of the left emoji in the sequence, appeared on the screen when the participants pressed the space bar. After looking at this fixation point for one second, the emoji sequence appeared on the screen. The numbers "1", "2", and "3", representing the possible answer options, simultaneously appeared on the screen. Participants submitted their answer by pressing the corresponding number on the keyboard (see Figure 4 for a visualization of an experimental trial).

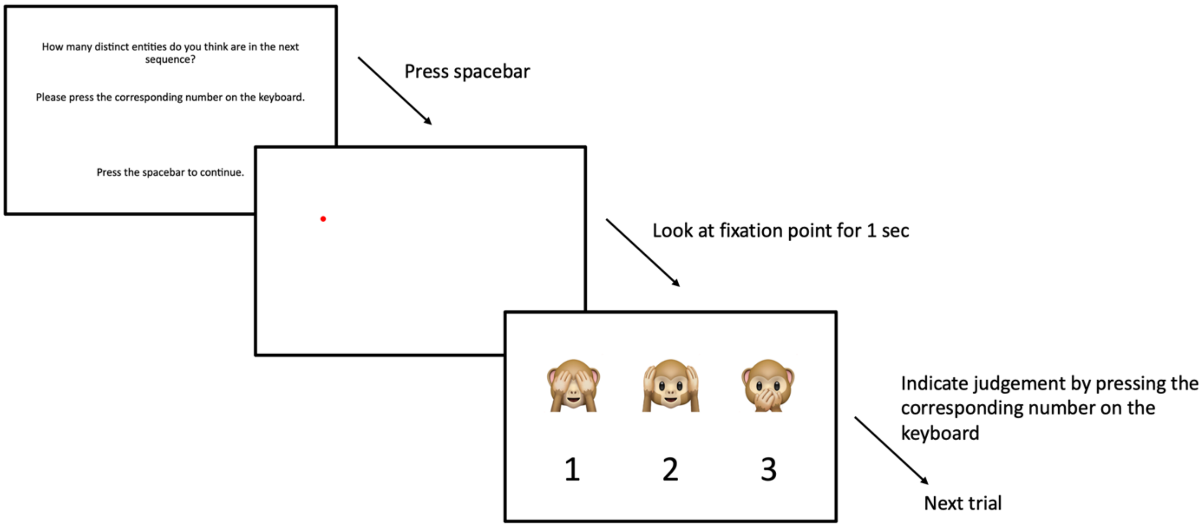

Figure 4. Visualization of an experimental trial.

When the participants finished the experiment, they were instructed to notify the experimenter. Participants were asked whether they noticed any patterns or if anything stood out to them during the experiment, after which they received an oral debriefing of the experiment. In total, the experimental sessions had a duration of approximately 40 minutes. The eye-tracking part took around 20-30 minutes.

3.7. Data analysis

My analysis focused on participants' eye movements while viewing the emoji sequences, the time they took to make judgements about the represented number of entities and the judgements itself. To prepare the dependent eye-movement measures (see Table 1), six areas of interest (AOIs) were created (see Figure 5). The AOIs required for the primary eye movement measures covered the emoji (400 x 400 pixels). Furthermore, the subsidiary eye movement measure also required AOIs covering each answer option (250 x 250 pixels). The calculation of each dependent eye movement measure can be seen in Table 1.

Table 1. Calculation of dependent eye movement measures.

Analysis	Dependent eye movement measure	Calculation
Primary analysis	Number of switches between emoji 1, 2, and 3.	Adding up the number of switches from AOI1-AOI2, AOI1-AOI3, AOI2-AOI3, AOI2-AOI1, AOI3-AOI1, AOI3-AOI2.
	Number of forward switches between emoji 1, 2, and 3	Adding up the number of switches from AOI1-AOI2, AOI1-AOI3, AOI2- AOI3.
	Number of backward switches between emoji 1, 2, and 3.	Adding up the number of switches from AOI2-AOI1, AOI3-AOI1, AOI3-AOI2.
	Number of switches from emoji 1 to 2.	Number of switches from AOI1 to AOI2.
	Number of switches from emoji 1 to 3.	Number of switches from AOI1 to AOI3.
	Number of switches from emoji 2 to 3.	Number of switches from AOI2 to AOI3.
	Number of switches from emoji 2 to 1.	Number of switches from AOI2 to AOI1.
	Number of switches from emoji 3 to 1.	Number of switches from AOI3 to AOI1.
	Number of switches from emoji 3 to 2.	Number of switches from AOI3 to AOI2.
	Total viewing time of emoji 1, 2 and 3.	Adding up the duration of all fixations on AOI1, AOI2, and AOI3.
Subsidiary analysis	Number of switches between emoji and answer options.	Adding up the number of switches from AOI1-AOI2, AOI1-AOI3, AOI2-AOI3, AOI2-AOI1, AOI3-AOI1, AOI3-AOI2,

	AOI1-AOI4, AOI1-AOI5, AOI1-AOI6, AOI2-AOI4, AOI2-AOI5, AOI2-AOI6, AOI3-AOI4, AOI3-AOI5, AOI3-AOI6, AOI4-AOI1, AOI4, AOI2, AOI4-AOI3, AOI4-AOI5, AOI4-AOI6, AOI5-AOI1, AOI5-AOI2, AOI5-AOI3, AOI5-AOI4, AOI5-AOI6, AOI6-AOI1, AOI6-AOI2, AOI6-AOI3, AOI6-AOI4, AOI6-AOI5.

Figure 5. Areas of interest covering the three emoji (blue) and the three answer options (red).

I used the software Fixation (Cozijn, 2006) to prepare the dependent eye movement measures for statistical analysis. Even though most fixations were automatically assigned to the AOIs, not all fixations could be assigned as a result of imperfect calibration or drifts leading to fixations landing too far above or below the AOIs. Unassigned fixations were assigned to the AOI of the next fixation when they had a duration of <100ms and/or could clearly be identified as a jump between the previous and next fixation. Unassigned fixations were excluded from eye movement analyses when they had a duration of >100ms and/or no clear jump between the previous and next assigned fixation.

Linear mixed models (LMMs) were used for each dependent eye movement measure and people's decision times. Each LMM included the categorical variables sequence nature (ordered vs. unordered) and scrambling (non-scrambled vs. scrambled) as fixed effects. These LMMs also included five categorical variables as fixed effects. Both categorical and colour similarity scores (retrieved from: Lichtenberg et al., In prep) were added as fixed effects. Additionally, an average ambiguity score for referential continuity was added as a fixed effect into the LMMS. This score was calculated by averaging ambiguity scores from Lichtenberg et al. (In prep) and the current study for each emoji sequence (see Appendix I for the calculation). Furthermore, comic reading and emoji experience (indicated by VLFI and ELF2 scores) were also included as continuous fixed effects into the LMMs. Lastly, participant number was added as a random effect into the LMMs, to account for individual differences between participants. A Bonferroni correction was used for post-hoc pairwise analyses. Eye movement measures categorized as primary analysis were run to investigate the effect of predictors on participants' eye movements solely focusing on the emoji, while the eye movement measure categorized as subsidiary analysis was run substantiate the analysis of participants' judgements (see Table 1).

For the judgements about the represented number of entities, three LMMs were hierarchically compared by looking at ML values. The first LMM only included the categorical variables sequence nature (ordered vs. unordered) and scrambling (non-scrambled vs. scrambled) as fixed effects. In the second LMM, continuous variables that were bound to the emoji sequence (categorical similarity, colour similarity and average ambiguity scores) were added as fixed effects. The third LMM also included continuous variables bound to the comprehender (comic reading and emoji experience) as fixed effects. All models included participant number as a random effect, to account for individual differences between participants. Moreover, to explore relations between participants' judgements and their eye movements, additional Pearson's correlation analyses were conducted between participants' judgements and dependent eye movement measures that generated significant effects in the LMMs.

Furthermore, all LMMs were built in the R programming environment (version 4.4.1), by using the 'lmer' function from the 'lme4' package (1.1.35.5) (Bates, Mächler, Bolker, & Walker, 2014). LMM visualizations were created using the 'sjPlot' package (version 2.8.16), which allows for visualizations specifically tailored for mixed-effects models such as interaction plots and predicted value plots (Lüdecke, 2024).

Lastly, additional Pearson's correlation analyses were conducted to explore relations between continuous predictors itself and between continuous predictors and dependent variables (DVs) of which the LMMs generated significant effects.

4. Results

4.1. Eye movement measures

In the next sections, eye movement analyses will be discussed. These sections will only discuss analyses that yielded significant effects (number of switches between emoji 1, 2 and 3; number of switches from emoji 1 to 3; total viewing time of emoji 1, 2 and 3). Analyses that did not detect significant effects can be viewed in Appendix II (number of forward switches between emoji 1, 2 and 3; number of backward switches between emoji 1, 2 and 3; number of switches from emoji 1 to 2; number of switches from emoji 2 to 3; number of switches from emoji 2 to 1; number of switches from emoji 3 to 1; number of switches from emoji 3 to 2).

4.1.1. Number of switches between emoji 1, 2 and 3

The fixed effects analysis of the number of switches between emoji 1, 2 and 3 revealed four main effects of sequence nature, order, categorical similarity score, and colour similarity score (see Table 2). It also showed an interaction between sequence nature and order. Participants, generally switched more between emoji 1, 2 and 3 for unordered than ordered sequences (see Figure 6). In addition, scrambled sequences led to more switches between emoji 1, 2 and 3 than non-scrambled sequences. Moreover, higher categorical similarity scores were associated with more switches between emoji 1, 2 and 3, while higher colour similarity scores were associated with fewer switches (see Figure 7).

Table 2. Fixed effects of the LMM predicting the number of switches between emoji 1, 2 and 3.

Predictor	**β**	**SD**	**t**	**p**
Intercept	4.83	8.11	0.60	.55
Sequence nature (unordered vs. ordered)	4.54	.79	5.75	< .001***
Order (scrambled vs. non-scrambled)	2.19	.85	2.56	.01*
Sequence nature * Order	-1.59	.67	-2.36	.02 *
Categorical similarity score	9.66	1.25	7.71	< .001***
Colour similarity score	-10.51	.92	-11.42	< .001***
Average ambiguity score	.01	1.04	.01	.99
VLFI score	.25	0.13	1.85	.07
ELF2 score	-.87	1.52	-0.57	.57

df.resid = 1,1164

Significance levels: ***p < .001, **p < .01, *p < .05

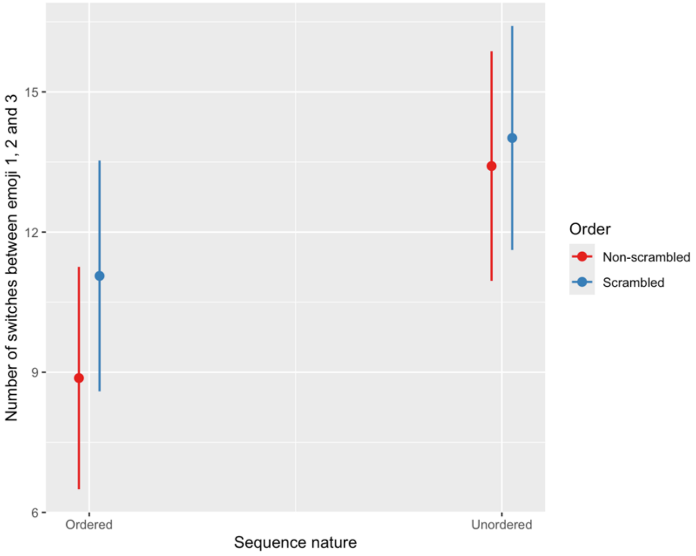

Figure 6. Number of switches between emoji 1, 2 and 3 for ordered and unordered sequences that were either non-scrambled or scrambled. Error bars represent 95% confidence intervals.

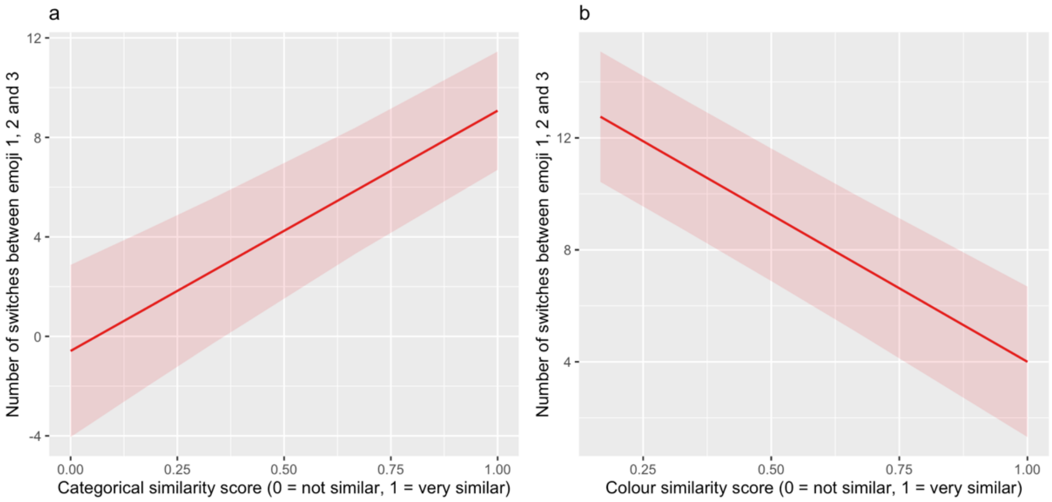

Figure 7. Number of switches between emoji 1, 2 and 3 predicted for a) categorical similarity scores of the emoji sequences, and b) colour similarity scores of the emoji sequences, displayed with 95% confidence intervals.

In further analyzing the interaction between sequence nature and order, post hoc tests revealed that participants made fewer switches between emoji 1, 2 and 3 if sequences were ordered compared to when they were unordered for both non-scrambled sequences (*Mdif* = -4.54, *SE* = .79, *p*<.0001) and scrambled sequences (*Mdif* = -2.95, *SE* = .87, *p* = .0042), see Figure 6. Participants also made fewer switches between emoji 1, 2 and 3 if they viewed ordered sequences in a non-scrambled order, compared to when they scrambled (*Mdif* = -2.19, *SE* = .85, *p* = .06). This effect was, however, only marginally significant.

The random effect of participant number showed a variance of 46.50 (*SD* = 6.81, indicating between-subject variability, and a residual variance of 24.95 (*SD* = 5.00, indicating within-subject variability).

4.1.2. Number of switches from emoji 1 to 3

The fixed effects analysis of the number of switches from emoji 1 to 3 revealed two main effects of order and average ambiguity score (see Table 3). It also showed an interaction between sequence nature and order. Participants, generally, switched more from emoji 1 to 3 when sequences were scrambled compared to when they were not scrambled (see Figure 8). In addition, higher average ambiguity scores were associated with more switches from emoji 1 to 3 (see Figure 9).

Table 3. Fixed effects of the LMM predicting the number of switches from emoji 1 to 3.

Predictor	**β**	**SD**	**t**	**p**
Intercept	-.19	.10	-1.96	.05*
Sequence nature (unordered vs. ordered)	.01	.03	.35	.72
Order (scrambled vs. non-scrambled)	.06	.02	3.23	.00**
Sequence nature * Order	-.06	.03	-2.23	.03*
Categorical similarity score	-.01	.05	-.23	.82
Colour similarity score	.00	.04	.03	0.97
Average ambiguity score	.09	.04	2.29	.02*
VLFI score	-.00	.00	-.33	.75
ELF2 score	.01	.01	1.12	.26

df.resid = 1,1064

Significance levels: ***p < .001, **p < .01, *p < .05

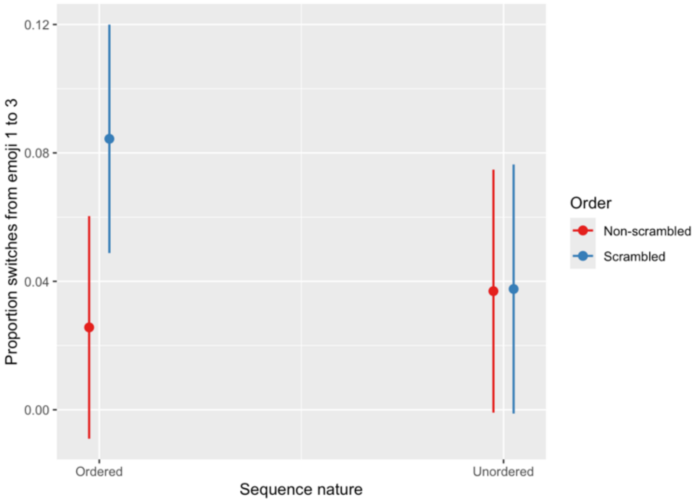

Figure 8. Proportion switches from emoji 1 to 3 for ordered and unordered sequences that were either non-scrambled or scrambled. Error bars represent 95% confidence intervals.

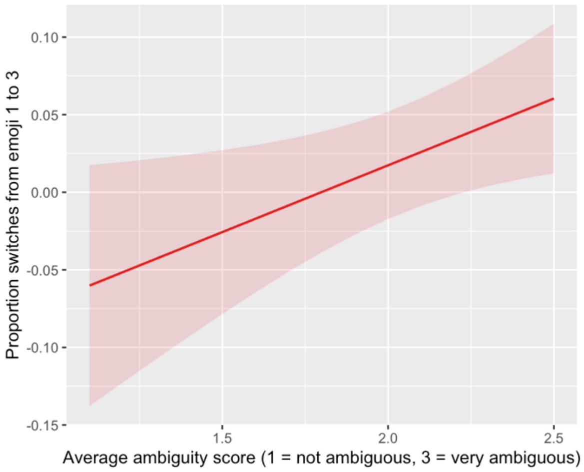

Figure 9. Proportion of switches from emoji 1 to 3 predicted for average ambiguity scores (1 = not ambiguous, 3 = very ambiguous), displayed with 95% confidence intervals.

In further analyzing the interaction between sequence nature and order, post hoc tests revealed that participants made fewer switches from emoji 1 to 3 if ordered sequences were non-scrambled compared to when they were scrambled (*Mdif* = -.06, *SE* = .02, *p* = .008), see Figure 8.

The random effect of participant number showed a variance of .00 (*SD* = .00, indicating between-subject variability, and a residual variance of .04 (*SD* = .21, indicating within-subject variability).

4.1.3. Total viewing time of emoji 1, 2 and 3

The fixed effects analysis of total viewing time of emoji 1, 2 and 3 revealed four main effects of sequence nature, categorical similarity score, colour similarity score, and average ambiguity (see Table 4). As depicted in Figure 10, participants looked longer at emoji 1, 2 and 3 in unordered sequences, compared to ordered sequences. In addition, higher categorical similarity scores were associated with longer viewing times of emoji 1, 2 and 3, while higher colour and average ambiguity scores were associated with shorter viewing times (see Figure 11).

Table 4. Fixed effects of the LMM predicting the total viewing time of emoji 1, 2 and 3.

Predictor	**β**	**SD**	**t**	**p**
Intercept	8574.51	4751.18	1.81	.08
Sequence nature (unordered vs. ordered)	2845.09	495.19	5.75	< .001***
Order (scrambled vs. non-scrambled)	-157.20	530.56	-.30	.77
Sequence nature * Order	5.98	420.35	.01	.99
Categorical similarity score	5474.19	786.15	6.96	< .001***
Colour similarity score	-5415.80	577.27	-9.38	< .001***
Average ambiguity score	-2274.77	651.74	-3.49	.001**
VLFI score	148.76	78.16	1.90	.06
ELF2 score	-807.40	881.13	-.92	.36

df.resid = 1,1164

Significance levels: ***p < .001, **p < .01, *p < .05

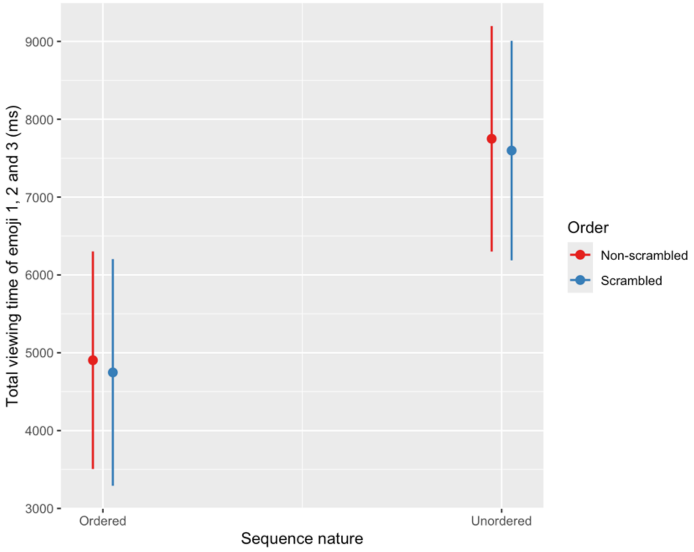

Figure 10. Total viewing time of emoji 1, 2 and 3 (ms) for ordered and unordered sequences that were either non-scrambled or scrambled. Error bars represent 95% CI.

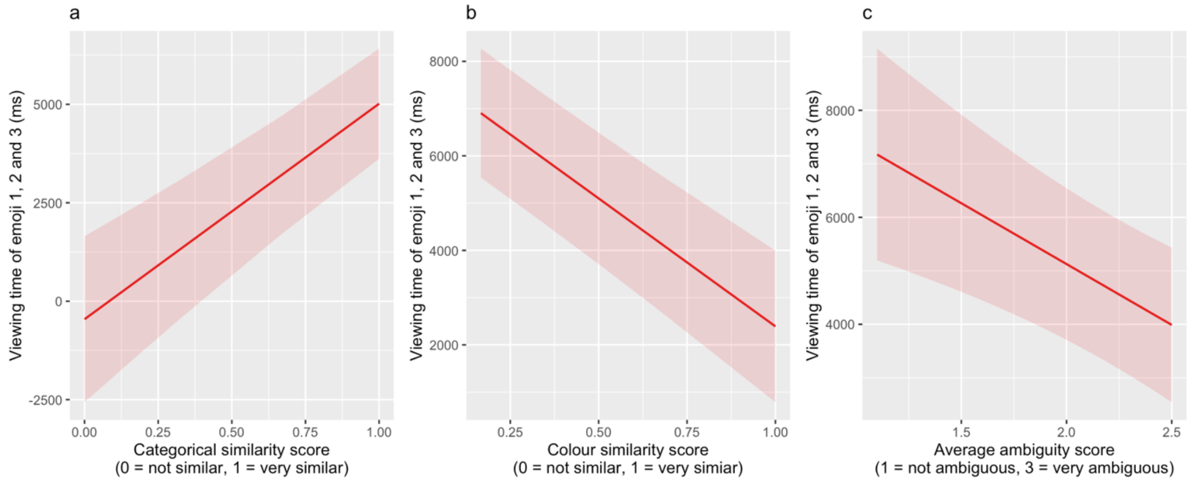

Figure 11. Viewing time of emoji 1, 2 and 3 (ms) predicted for a) categorical similarity scores of the emoji sequences, b) colour similarity scores of the emoji sequences, and c) average ambiguity scores of the emoji sequences, displayed with 95% confidence intervals.

The random effect of participant number showed a variance of 15569143 (*SD* = 3946, indicating between-subject variability, and a residual variance of 9825217 (*SD* = 3135, indicating within-subject variability).

4.2. Decision time

The fixed effects analysis of decision time revealed four main effects of sequence nature, categorical similarity score, colour similarity score, and average ambiguity score (see Table 5). As depicted in Figure 12, participants, generally, took more time to make judgements about unordered sequences, compared to ordered sequences. Furthermore, higher categorical similarity scores were associated with longer decision times, while higher colour and average ambiguity scores were associated with shorter decision times (see Figure 13).

Table 5. Fixed effects of the LMM predicting the decision time.

Predictor	**β**	**SD**	**t**	**p**
Intercept	6656.43	5606.70	1.19	.24
Sequence nature (unordered vs. ordered)	3025.52	590.90	5.12	< .001***
Order (scrambled vs. non-scrambled)	404.12	632.10	.64	.52
Sequence nature * Order	-226.43	501.44	-.45	.65
Categorical similarity score	6971.57	938.01	7.43	< .001***
Colour similarity score	-6325.25	688.86	-9.18	< .001***

Predictor	**β**	**SD**	**t**	**p**
Average ambiguity score	-1537.10	777.23	-1.98	.05*
VLFI score	162.59	92.09	1.77	.09
ELF2 score	-772.05	1038.00	-.74	.46

df.resid = 1,1164

Significance levels: ***p < .001, **p < .01, *p < .05

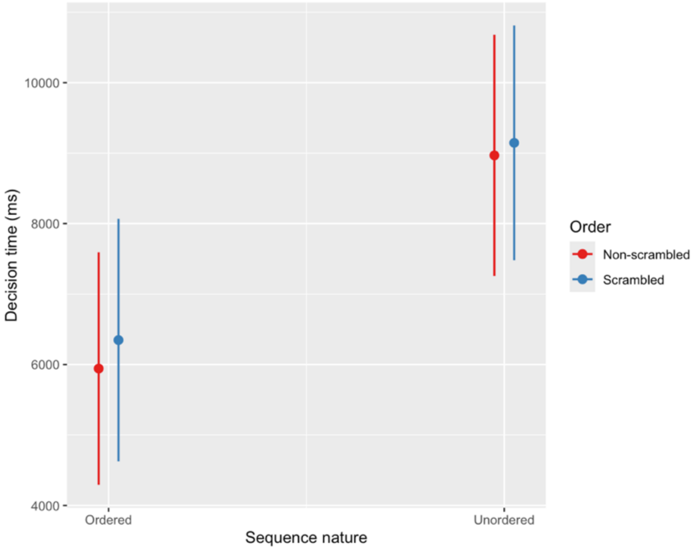

Figure 12. Decision time (ms) for ordered and unordered sequences that were either non-scrambled or scrambled. Error bars represent 95% CI.

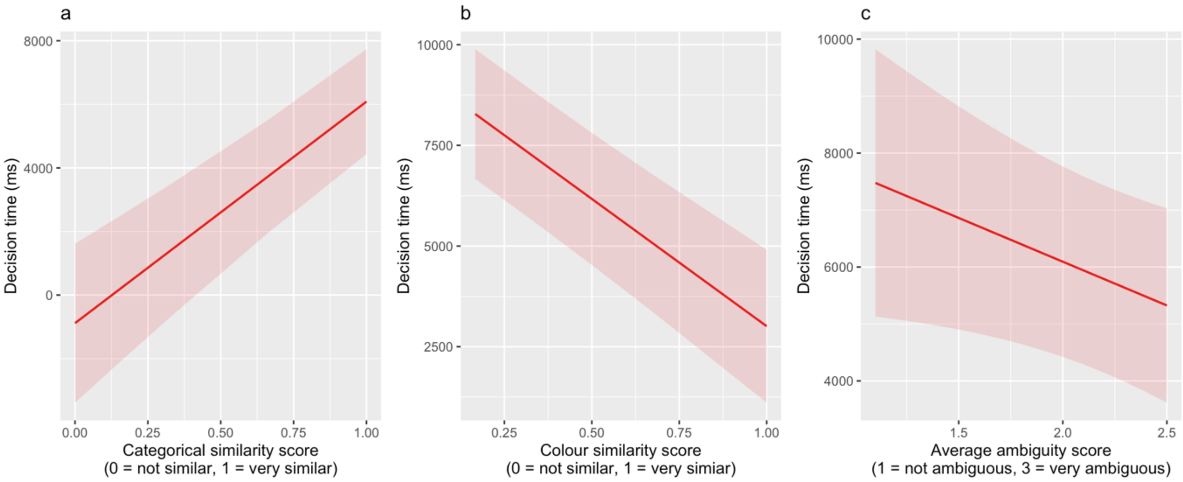

Figure 13. Decision time (ms) predicted for a) categorical similarity scores of the emoji sequences, b) colour similarity scores of the emoji sequences, and c) average ambiguity scores of the emoji sequences, displayed with 95% confidence intervals.

The random effect of participant number showed a variance of 21587685 (*SD* = 4646, indicating between-subject variability, and a residual variance of 13990972 (*SD* = 3740, indicating within-subject variability).

4.3. Judgements

4.3.1. Judgements

To hierarchically compare the three LMMs that were run to analyze participants' judgements about their fit statistics were compared (see Table 6). Even though Model 1 has the highest log-likelihood and Model 3 has the lowest deviance, Model 2 has the lowest AIC and BIC values, suggesting that Model 2 provides the best balance of model complexity and fit. For that reason, the output of this model will be discussed in the following sections. The output of the first and the third model can be viewed in Appendix III.

Table 6. Fit statistics for model 1, 2 and 3 predicting judgements.

Model	**AIC**	**BIC**	**Log-likelihood**	**Deviance**
Model 1	5950.6	5985.3	-2969.3	5938.6
Model 2	5540.8	5592.8	-2761.4	5522.8
Model 3	5544.2	5607.9	-2761.1	5522.2

The fixed effects analysis of the second model predicting judgements revealed four main effects of sequence nature, categorical similarity score, colour similarity score, and average ambiguity score (see Table 7). As depicted in Figure 14, participants, were more likely to construct continuity for ordered sequences compared to unordered sequences. Moreover, participants were more likely to construct continuity as categorical and colour similarity scores went up. In contrast, they were less likely to construct continuity as average ambiguity ratings went up (see Figure 15).

Table 7. Fixed effects of the second LMM predicting judgements.

Predictor	**β**	**SD**	**t**	**p**
Intercept	2.58	.12	21.70	<.001***
Sequence nature (unordered vs. ordered)	.46	.05	9.88	<.001***
Order (scrambled vs. non-scrambled)	.01	.04	.14	.89
Sequence nature * Order	-.01	.06	-.22	.83
Categorical similarity score	-.43	.06	-6.83	<.001***
Colour similarity score	-1.01	.06	-15.75	<.001***

Predictor	**β**	**SD**	**t**	**p**
Average ambiguity score	.12	.05	2.38	.02*

df.resid = 1,2391

Significance levels: ***p < .001, **p < .01, *p < .05

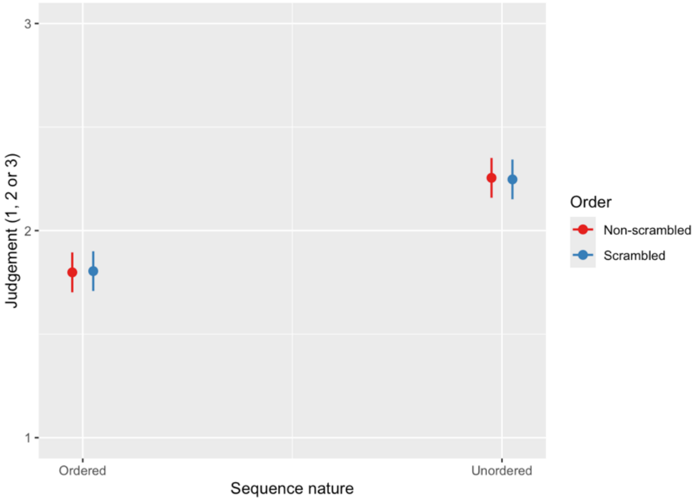

Figure 14. Judgements (1 for one entity, 2 for two entities, and 3 for three entities) for ordered and unordered sequences that were either non-scrambled or scrambled. Error bars represent 95% confidence intervals.

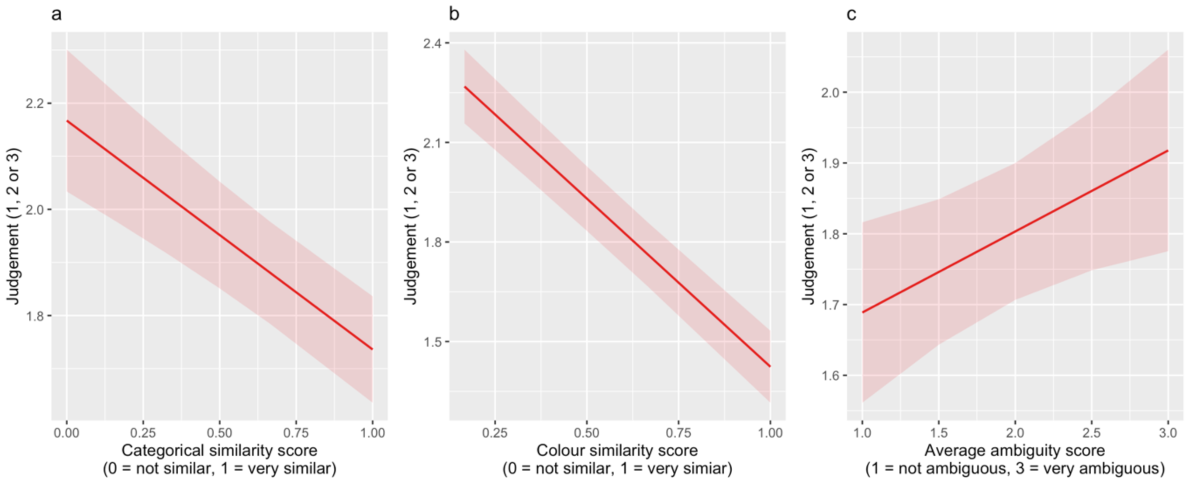

Figure 15. Judgements predicted for a) categorical similarity scores of the emoji sequences, b) colour similarity scores of the emoji sequences, and c) average ambiguity scores of the emoji sequences, displayed with 95% confidence intervals.

The random effect of participant number showed a variance of .06 (*SD* = .24, indicating between-subject variability, and a residual variance of .57 (*SD* = .75, indicating within-subject variability).

Exploratory Pearson’s correlations were then performed between participants’ judgements and dependent eye movement measures that generated significant effects in the LMMs. A negative correlation appeared between participants’ judgements and the number of switches between emoji 1, 2 and 3, *r*(158) = -.16, *p* =.048 . This indicated that constructing continuity between emoji in sequences was associated with fewer switches between the emoji in these sequences (see Figure 16).

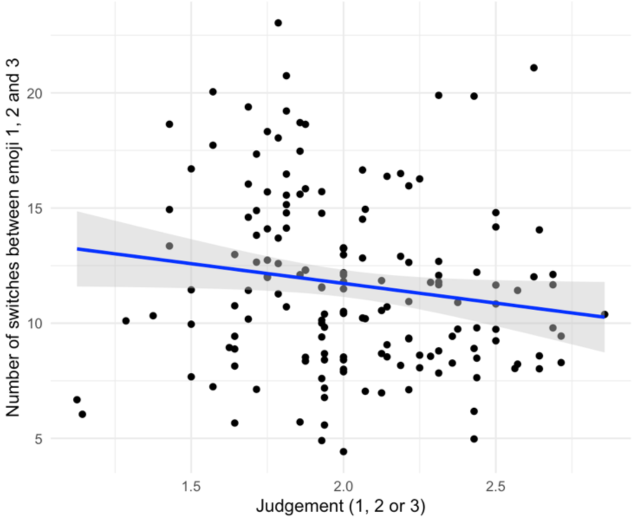

Figure 16. Judgements (averaged across participants) correlated with number of switches between emoji 1, 2 and 3 (averaged across participants). Grey area displays 95% confidence intervals.

4.3.2. Number of switches between emoji and answer options

The fixed effects analysis of the number of switches between both the emoji and answer options revealed four main effects of sequence nature, order, categorical similarity score and colour similarity score (see Table 8). It also showed an interaction between sequence nature and order. As depicted in Figure 17, participants, generally made more switches between the emoji and answer options when they viewed unordered sequences compared to ordered sequences. In addition, scrambled sequences led to more switches than non-scrambled sequences. Lastly, higher categorical similarity scores were associated with more switches, while higher colour similarity scores were associated with fewer switches (see Figure 18).

Table 8. Fixed effects of the LMM predicting the number of switches between emoji and answer options.

Predictor	**β**	**SD**	**t**	**p**
Intercept	7.22	9.05	.78	.43
Sequence nature (unordered vs. ordered)	5.74	.89	6.42	< .001***
Order (scrambled vs. non-scrambled)	4.07	.97	4.22	< .001***
Sequence nature * Order	-2.12	.76	-2.79	.005**
Categorical similarity score	8.18	1.42	5.76	< .001***
Colour similarity score	-11.385	1.04	-10.92	< .001***
Average ambiguity score	-.17	1.18	-.14	.89
VLFI score	.24	.15	1.60	.12
ELF2 score	-.75	1.69	-.44	.66

df.resid = 1,1164

Significance levels: ***p < .001, **p < .01, *p < .05

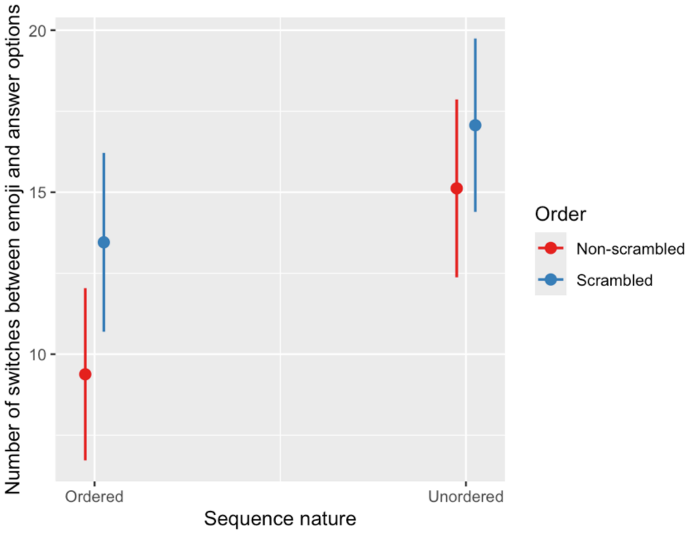

Figure 17. Number of switches between emoji and answer options for ordered and unordered sequences that were either non-scrambled or scrambled. Error bars represent 95% confidence intervals.

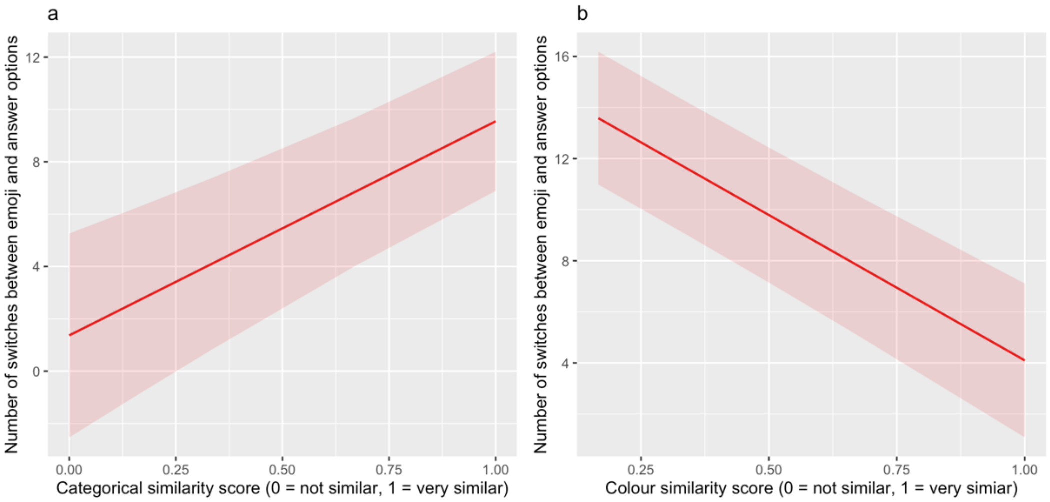

Figure 18. Number of switches between emoji and answer options predicted for a) categorical similarity scores of the emoji sequences and b) colour similarity scores, displayed with 95% CI.

In further analyzing the interaction between sequence nature and order, post hoc tests revealed that participants made fewer switches if non-scrambled sequences were ordered compared to when they were unordered (*Mdif* = -5.74, *SE* = .89, *p* < .001), see Figure 17. It was also found that participants made fewer switches if they viewed ordered sequences in a non-scrambled order, compared to when they scrambled (*Mdif* = -7.69, *SE* = 1.01, *p* < .001).

The random effect of participant number showed a variance of 57.64 (*SD* = 7.59, indicating between-subject variability, and a residual variance of 32.04 (*SD* = 5.66, indicating within-subject variability).

4.4. Exploratory correlation analyses

Exploratory Pearson's correlations firstly revealed a positive correlation between participants' general visual fluency and their emoji fluency (*r*(158) = .21, *p* = .007). This indicated that higher visual fluency was associated with higher emoji fluency (see Figure 19).

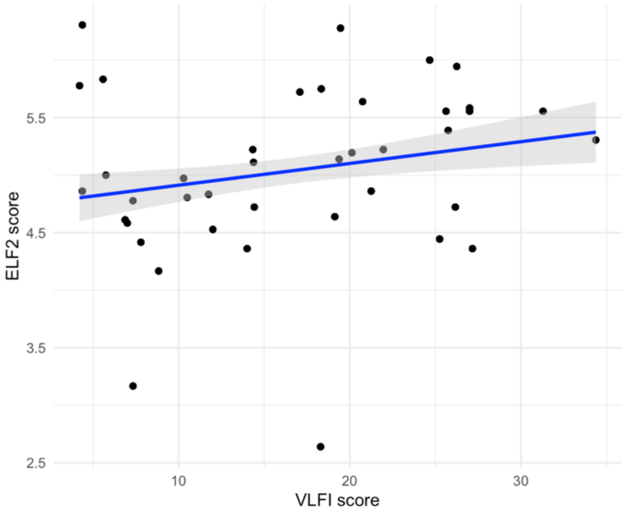

Figure 19. VLFI scores correlated with ELF2 scores. Grey area displays 95% confidence intervals.

Furthermore, exploratory correlations between continuous predictors and DVs that generated significant effects in the LMMs revealed three negative correlations (see Table 9). As can be seen in Figure 20, higher emoji fluency was associated with fewer switches between emoji 1, 2 and 3, shorter decision times and fewer switches between emoji and answer options.

Table 9. Significant Pearson's correlations between Emoji fluency and DVs that generated significant effects in the LMMs.

Continuous predictor	DV	r	p
Emoji fluency (ELF2 score)	Number of switches between emoji 1, 2 and 3	-.17	.03*
Emoji fluency (ELF2 score)	Decision time	-.21	.007**
Emoji fluency (ELF2 score)	Number of switches between emoji and answer options	-.18	.02*

df = 158

Significance levels: ***p < .001, **p < .01, *p < .05

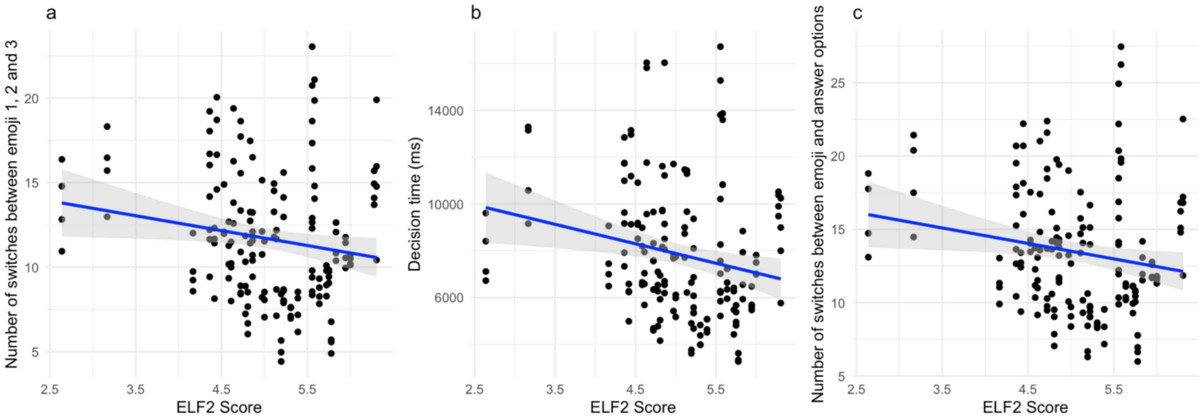

Figure 20. ELF2 scores, correlated with a) number of switches between emoji 1, 2 and 3, b) decision times (ms), and c) number of switches between emoji and answer options. All averaged across participants. Grey areas display 95% confidence intervals.

5. Discussion

With this study, I measured participants' eye movements while they viewed ordered and unordered sequences of three emoji that were either scrambled or not to investigate why emoji sequences like the "three-wise-monkey" can yield different interpretations about their distinct referential entities. Overall, I found that the interplay between visual features and individual experience influences people to construe continuity between emoji in sequences or not. Furthermore, the results can be interpreted as aligning with findings related to the processing of linguistic co-reference, thereby suggesting at more modality general behaviors operating on referential sequences.

As expected, participants judged the number of referential entities more quickly for ordered emoji sequences than unordered emoji sequences, and these ordered emoji were judged to have more continuity (i.e. fewer entities) than unordered emoji sequences, which replicates Lichtenberg et al. (In prep) and confirms H1a. These findings were further substantiated by participants' eye movements towards the answer options. The recognition of continuity between entities across image-units often guides people to capture that the sequence is bound to be displayed in a specific order (e.g. the journey of a Christmas tree) (Cohn, 2020a). On the other hand, people do not have to grasp this sequential nature while viewing emoji sequences that do not adhere a fixed order, allowing for more sequential flexibility. For that reason, it is conceivable that people are more confident to construe co-reference between emoji in ordered sequences than unordered sequences.

Confirming H1b, this also manifested in participants' eye movements as participants were less likely to switch between emoji and spent less time looking at emoji in ordered sequences compared to unordered sequences. This is in line with prior work from Foulsham et al. (2016), in which comic panels presented in the original order induced fewer regressive eye movements and shorter fixations. These findings are also analogous to linguistic eye-tracking research showing that continuity between anaphora induced shorter fixations, shorter reading times and fewer regressive switches (Rayner et al., 2006; Cook, 2005, cited in Rayner, 2006). Taken together, they may suggest domain-general comprehension processes for referential sequencing.

To further validate the differences between ordered and unordered sequencing, I manipulated the sequences to also have a scrambled version in which I changed the positions of the emoji. As predicted, I found that scrambling the positions of the emoji affected participants' eye movements only for ordered sequences, as it induced more switches between emoji. However, contrary to Lichtenberg et al. (In prep), scrambling did not affect participant's judgements about the number of distinct entities represented in ordered sequences, as no effects regarding scrambling were detected in participants' judgements, through which H3ab is only partially confirmed. These inconsistent findings might be attributed to the fact that participants in the current study were instructed to take their time to carefully judge the sequences, while participants in Lichtenberg et al. (In prep) were encouraged to respond quickly. Even though people's eye movements suggest that changing the positions of the emoji in ordered sequences manifests in processing costs, the absence of time pressure could have encouraged participants to put in the effort to make sense of these scrambled ordered sequences anyway.

In addition, consistent with H2a, emoji sequences that exhibited greater similarity in terms of colour or superordinate semantic category (e.g. all emoji being monkeys) were also associated with judgements of fewer entities. These findings align with theoretical work proposing that such visual similarities can sponsor the construction of continuity across image-units (Klomberg et al., 2023), and replicate results from Lichtenberg et al. (In prep). H2a, was, however only partially confirmed as people were only faster to make decisions about sequences with emoji that had similar colours. In contrast, participants took more time to judge sequences that included emoji that were more similar in terms of superordinate semantic category. Moreover, this opposite pattern persisted in people's eye movements as greater colour similarity induced fewer switches between emoji, shorter viewing times and shorter decision times, while higher categorical similarity manifested in more switches between emoji, longer viewing times and longer decision times.

These unpredicted findings related to greater categorical similarity in sequences may be explained by the presence of alignable differences (e.g. the colour difference between two cars) rather than non-alignable differences (i.e. the difference between a car and a table) between the emoji (Gentner & Markman, 1994; Markman & Gentner, 1996). As alignable differences allow for a direct comparison between units, making it easier for people to identify them, it is posited that they are more important than non-alignable differences when people make similarity judgements. Therefore, it is conceivable that participants' decision times and eye movements reflected processing mechanisms underlying the evaluation of these alignable difference between emoji (e.g. all faces with different colours) that were not manifested in their conscious judgements.

Moreover, as predicted in H4a, sequences that were more ambiguous regarding referential continuity (i.e. there was less agreement between participants about the number of distinct entities) were associated with judgements of more entities, aligning with Lichtenberg et al. (In prep). Consistent with H4b, this was also shown in in participants' eye movements as participants made more switches from the first to the third emoji. This is similar to linguistic eye-tracking research that indicated that ambiguous anaphors induced more switching eye movements to antecedents compared to unambiguous anaphors (Spivey-Knowlton & Tanenhaus, 2015), suggesting that more ambiguous sequences require more cognitive resources to process.

However, the referential ambiguity of the emoji sequences also yielded unexpected results, through which H4a, and H4b could only partially be confirmed. Inconsistent with the findings of Lichtenberg et al. (In prep), people judged more ambiguous sequences faster than less ambiguous sequences. In addition, more ambiguous sequences were associated with shorter viewing times of the emoji, contrasting linguistic eye-tracking research showing that more ambiguous anaphors induced longer reading times (Spivey-Knowlton & Tanenhaus, 2015). As the current results opposite prior findings, future research is encouraged to clarify these discrepancies by investigating how various properties of emoji sequences relate to their referential ambiguity and impact people's decision times and emoji viewing times.

Furthermore, no effects were found on participants' overall regressions. This is not in line with verbal eye-tracking studies which indicated that discontinuity or ambiguity between anaphora only increased regressive eye movements (Rayner, et al., 2006; Cook, 2005, cited in Rayner, 2006; Spivey-Knowlton & Tanenhaus, 2015), and contrasts H1b, H2b, and H4b. As the current study included more referential units (three emoji) than these verbal studies (two: anaphor and antecedent), the possible switches that participants could make between them was increased. This increase may contribute to the detection of making more switches in general, rather than an increase in regressive switches exclusively.

Another explanation could be found in the discrepancy in distances between the referential units in the linguistic studies (i.e. anaphor and referent with words in between) and the referential units in the current study (i.e. no distance between the emoji). As prior work showed that greater distances between referential units manifest in processing costs across modalities (Gibson, 2000; Cohn et al., 2024), it is conceivable that the proximity of the emoji placement in this study did not require people to make extra backward switches. Future research could, therefore, investigate longer sequences of emoji and manipulate distances between co-referential emoji.

Besides properties of the emoji, as predicted in H5ab, participants' prior experience with emoji affected their eye movements and decision times. Participants with greater emoji experience made fewer switches between the emoji in sequences before submitting their judgements and made their judgements faster, implying faster access to the principles underlying continuity in emoji sequences. This is in line with findings that more emoji fluent people responded faster and were more likely to construct continuity between them in a sequence (Lichtenberg et al., In prep). Furthermore, it aligns with Cohn and Kutas (2017), who also found that a specific type of visual fluency (Manga readership) can modulate visual narrative comprehension.

Nevertheless, in line with prior work (Lichtenberg et al., In prep), this proficiency was limited to the experience with emoji specifically, not to a general proficiency with visual narratives. No relation was found between participants' overall visual language expertise and their eye movementss and judgements. This may imply that people's specific visual fluency, rather than their general visual fluency, makes the principles underlying continuity within specific visual sequences (e.g. emoji sequences) more accessible. It should, however, be noted that effects regarding people's specific emoji experience were only detected in exploratory correlation analyses. For that reason, future research can investigate how aspects of specific fluencies can modulate the comprehension across specific types of visual sequences.

6. Conclusion

To conclude, this study showed that, despite their small and simple nature, people can interpret emoji differently when they are presented in sequences and that this is also reflected in their eye movements. This challenges the idea that emoji can be understood universally (Ai et al., 2017; Evans, 2017). Certain properties of emoji sequences, such as categorical similarity, colour similarity, as well as consensus on the number of distinct referential entities, can motivate people to construct continuity and a sense of order in the sequences.

In addition, these results further substantiated the importance of specific fluencies in sequential image processing (Cohn & Kutas, 2017; Lichtenberg et al., In prep). The notion that emoji experience was found to modulate referential processing in sequences of small and simple images questions the assumption that visual sequences are universally accessible. It also illustrates the need for future research on how specific visual fluencies can modulate sequential image processing.

Moreover, this study found similar results as those examining eye movements during verbal co-reference in sentences (i.e. continuity and ambiguity) (Rayner, et al., 2006; Cook, 2005, cited in Rayner, 2006; Spivey-Knowlton & Tanenhaus, 2015), indicating the potential for similar constraints on affecting eye movements during verbal and visual co-reference. This suggests that patterns related to the processing of referential sequences may persist across domains. The study, therefore, aligns with findings that, despite having different representations, the sequencing of referential dependencies relies on domain-general constraints rather than domain-specific constraints (Cohn et al., 2024).

Lastly, the findings of this study can explain why the “three-wise-monkey” emoji in Figure 1 have an ambiguous interpretation (Wilde, 2019). The emoji in this sequence belong to the same superordinate semantic category (all monkeys) and have similar colours. These results show that, in line with theoretical work on visual co-reference (Klomberg et al., 2023), such features cue comprehenders to construe continuity between the emoji (82.5% of the participants in this study. Nevertheless, some comprehenders still perceive the sequence as representing three distinct monkeys (17.5% of the participants in this study).

The results also higlighted the importance of people’s prior exposure ot emoji sequences. Both people’s eye movements and their decision times suggested that more emoji experience provided an advantage in accessing the principles underlying the construction of continuity (possibly explaining why distinct interpretations of the “three-wise-monkey” emoji were not maintained equally here, given the emoji fluent study sample). This aligns with prior work indicating that more emoji experience was associated with the construal of more continuity between emoji in sequences (Lichtenberg et al., In prep). For that reason, it may be conceivable that people who fail to construct continuity between the monkeys in the “three-wise-monkey” lack sufficient experience with emoji, making the underlying princples of continuity less accessible.

Finally, by emphasizing the interplay between visual cues and individuals' experience, this study contributed to the understanding of continuity construal between emoji in sequences like the "three-wise-monkey" emoji. It also supports the idea that factors constraining co-reference persist across domains, suggesting that multiple modalities share mechanisms underlying the processing of referential sequencing.

References

Ai, W., Lu, X., Liu, X., Wang, N., Huang, G., & Mei, Q. (2017). Untangling emoji popularity through semantic embeddings. *Proceedings of the International AAAI Conference on Web and Social Media*, *11*(1).

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Bishop, A. (1977). Is a Picture Worth a Thousand Words? *Mathematics Teaching*, *81*, 32-35.

Bornens, M. T. (1990). Problems brought about by "reading" a sequence of pictures. *Journal of Experimental Child Psychology*, *49*(2), 189-226.

Byram, M. L., & Garfoth, C. (1980). Research and testing non-formal education materials: A multi-media extension project in Botswana. *Educational Broadcasting International*, *13*(4), 190-194.

Chomsky, N. (1980). On Binding. *Linguistic Inquiry*, *11*(1), 1-46.

Coderre, E. L., & Cohn, N. (2023). Individual differences in the neural dynamics of visual narrative comprehension: The effects of proficiency and age of acquisition. *Psychonomic Bulletin & Review*, *31*(1), 89-103.

Cohn, N. (2019). Visual narratives and the mind: Comprehension, cognition, and learning. *Psychology of learning and motivation*, *70*, 97-127.

Cohn, N. (2020a). Visual narrative comprehension: Universal or not? *Psychonomic Bulletin & Review*, *27*(2), 266-285.

Cohn, N. (2020b). *Who understands comics?: Questioning the universality of visual language comprehension*. Bloomsbury Publishing.

Cohn, N., Engelen, J., & Schilperoord, J. (2019). The grammar of emoji? Constraints on communicative pictorial sequencing. *Cognitive research: principles and implications*, *4*(1), 1-18.

Cohn, N., & Kutas, M. (2017). What's your neural function, visual narrative conjunction? Grammar, meaning, and fluency in sequential image processing. *Cognitive research: principles and implications*, *2*, 1-13.

Cohn, N., van Middelaar, L., Foulsham, T., & Schilperoord, J. (2024). Anaphoric distance dependencies in visual narrative structure and processing. *Cognitive Psychology*, *149*, 101639.

Cook, B. L. (1980). Picture communication in the Papua New Guinea. *Educational Broadcasting International*, *13*(2), 78-83.

Cozijn, R. (2006). Het gebruik van oogbewegingen in leesonderzoek [The use of eye movements in reading research]. *Tijdschrift voor Taalbeheersing*, *28*, 1270-1276.

Dürscheid, C., & Haralambous, Y. (2021). Emojis are everywhere. How emojis conquer new contexts. *Grapholinguistics and its applications*, *4*, 501-512.

Evans, V. (2017). *The emoji code: How smiley faces, love hearts and thumbs up are changing the way we communicate*. Michael O'Mara Books.

Foulsham, T., Wybrow, D., & Cohn, N. (2016). Reading without words: Eye movements in the comprehension of comic strips. *Applied Cognitive Psychology*, *30*(4), 566-579.

Fussell, D., & Haaland, A. (1978). Communicating with Pictures in Nepal: Results of Practical Study Used in Visual Education. *Educational Broadcasting International*, *11*(1), 25-31.

Gawne, L. (2016). A sketch grammar of Lamjung Yolmo. *Asia-Pacific Linguistics, School of Culture, History and Language, College of Asia and the Pacific, The Australian National University.*

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*.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In *Image, language, brain* (pp. 95-126). MA: MIT Press.

Gordon, P. C., & Hendrick, R. (1997). Intuitive knowledge of linguistic co-reference. *Cognition*, *62*(3), 325-370.

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, *48*, 163-189.

Hutson, J. P., Magliano, J. P., & Loschky, L. C. (2018). Understanding moment-to-moment processing of visual narratives. *Cognitive Science*, *42*(8), 2999-3033.

Jackendoff, R. (1983). *Semantics and cognition*. MIT Press.

Jackendoff, R. (2010). *Meaning and the Lexicon: The Parallel Architecture 1975-2010.* Oxford University Press.

Karabanov, A., Bosch, P., & König, P. (2007). Eye tracking as a tool to investigate the comprehension of referential expressions. *Roots. Linguistics in Search of its Evidential Base*, 207-226.

Klomberg, B., Hacımusaoğlu, I., Lichtenberg, L. D., Schilperoord, J., & Cohn, N. (2023). Continuity, Co-reference, and Inference in Visual Sequencing. *Glossa: a journal of general linguistics*, *8*(1).

Laubrock, J., Hohenstein, S., & Kümmerer, M. (2018). Attention to comics: Cognitive processing during the reading of graphic literature. In *Empirical comics research* (pp. 239-263). Routledge.

Lichtenberg, L. D., Hacimusaoğlu, I., Klomberg, B., Schilperoord, J., & Cohn, N. (In prep). *A closer look to the monkey emoji debate. Assessing the continuity constraint using emoji sequences.*

Lüdecke, M. D. (2024). *Package 'sjPlot'.*

Markman, A. B., & Gentner, D. (1996). Commonalities and differences in similarity comparisons. *Memory & Cognition*, *24*(2), 235-249.

McCulloch, G., & Gawne, L. (2018). *Emoji grammar as beat gestures*. 1-4.

Monti, J., Sangati, F., Chiuasaroli, F., Benjamin, M., & Mansour, S. (2016). *Emojitalianobot and EmojiWorldBot. 211.*

Nakazawa, J. (2002). Analysis of manga (comic) reading processes: Manga literacy and eye movement during Manga reading. *Manga Kenkyuu (Manga Studies)*, *5*(39).

Nakazawa, J., & Shwalb, D. W. (2012). Japan and the US comparison of university students' Manga reading literacy. *Paper presented at the Proceedings of Annual Conference of 54th Japanese Association of Educational Psychology*.

Paggio, P., & Tse, A. P. P. (2022). Are Emoji Processed Like Words? An Eye-Tracking Study. Cognitive Science. *Are Emoji Processed Like Words? An Eye-Tracking Study*, *46*(2), e13099.

Pohl, H., Domin, C., & Rohs, M. (2017). Beyond just text: Semantic emoji similarity modeling to support expressive communication. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *24*(1), 1-42.

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific studies of reading*, *10*(3), 241-255.

Robus, C. M., Hand, C. J., Filik, R., & Pitchford, M. (2020). Investigating effects of emoji on neutral narrative text: Evidence from eye movements and perceived emotional valence. *Computers in human Behavior*, *109*, 106361.

San Roque, L., Gawne, L., Hoenigman, D., Colleen Miller, J., Rumsey, A., Spronck, S., Caroll, A., & Evans, N. (2012). Getting the Story Straight: Language Fieldwork Using a Narrative Problem-Solving Task. *Language documentation and conservation*, *6*, 135-174.

Sanders, T. J. M., & Gernsbacher, M. A. (2004). Accessibility in Text and Discourse Processing. *Discourse Processes*, *37*(2), 79-89.

Scall, R. (2015). Emoji as language and their place outside American copyright law. *NYU J.*

Intell. Prop. & Ent. L., *5*(2), 381-405.

Scheffler, T., Brandt, L., de la Fuente, M., & Nenchev, I. (2022). The processing of emoji-word substitutions: A self-paced-reading study. *Computers in Human Behavior*, *127*, 107076.

Spivey-Knowlton, M., & Tanenhaus, M. (2015). Referential context and syntactic ambiguity resolution. *Tanenhaus*, 415-439.

Tatman, R. (2018). *Are emoji sequences as informative as text?* Making noise & hearing things. https://makingnoiseandhearingthings.com/2018/07/07/are-emoji-sequences-as-informative-as-text/

Trabasso, T., & Nickels, M. (1992). The development of goal plans of action in the narration of a picture story. *Discourse Processes*, *15*, 249-275.

Tseng, C. I., Laubrock, J., & Pfaeging, J. (2018). Character developments in comics and graphic novels: A systematic analytical scheme. In *Empirical Comics Research* (pp. 154-175). Routledge.

Weissman, B. (2019). *Emojis in sentence processing: An electrophysiological approach*. the Companion Proceedings of The 2019 World Wide Web Conference, San Francisco, CA.

Weissman, B., Engelen, J., Baas, E., & Cohn, N. (2023). The lexicon of emoji? Conventionality modulates processing of emoji. *Cognitive Science*, *47*(4), e13275

Wilde, R. (2019). *Emoticons, Kaomoji, and Emoji: Vol. The Elephant in the Room of Emoji Research: Or, Pictoriality, to what Extent? (pp. 171-196)*. Routledge.

Appendices

Appendix I – Calculation average ambiguity scores

Similarly to Lichtenberg et al. (In prep), I calculated ‘emoji ambiguity scores’ for the number of distinct referential entities represented in each emoji sequence. These scores serve as an indicator for the degree to which participants agreed with each other in their judgements. Similar to the ambiguity Lichtenberg et al. (In prep), the ambiguity scores for this study were calculated using the following formula:

$$\frac{(100-|(\%1\ entity-\%2\ entities)|)+(100-|(\%1\ entity-\%\ 3\ entities)|)+(100-|(\%\ 2\ entities-\%\ 3\ entities)|)}{300} / \ 0.33$$

Ambiguity scores ranged from 1 (no ambiguity, i.e. participants had the same intuitions about the number of distinct entities) to 3 (very ambiguous, i.e. participants had different intuitions about the number of distinct entities.

In order to make the referential ambiguity of the emoji score a more robust continuous predictor, I decided to average the ambiguity scores of Lichtenberg et al. (In prep) with the ambiguity scores of the current study. These average ambiguity scores were included in the analyses.

Appendix II – Eye movement measure analyses without significant effects

Number of forward switches between emoji 1, 2 and 3

The fixed analysis of number of forward switches between emoji 1, 2 and 3 did not reveal any significant main effect or a significant interaction (see Table 10). The random effect of participant number showed a variance of .00 (*SD* = .00, indicating between-subject variability, and a residual variance of .25 (*SD* = .50, indicating within-subject variability).

Table 10. Fixed effects of the LMM predicting the number of forward switches between emoji 1, 2 and 3.

Predictor	**β**	**SD**	**t**	**p**
Intercept	.35	.24	1.46	.15
Sequence nature (unordered vs. ordered)	-.02	.08	-.26	.80
Order (scrambled vs. non-scrambled)	-.03	.04	-.59	.55
Sequence nature * Order	.01	.06	.14	.89
Categorical similarity score	.21	.12	1.72	.09
Colour similarity score	-.01	.09	-.09	.93
Average ambiguity score	.02	.09	.24	.81
VLFI score	-.00	.00	-.84	.40
ELF2 score	-.01	.02	-.37	.71

df.resid = 1,1064

Significance levels: ***p < .001, **p < .01, *p < .05

Number of backward switches between emoji 1, 2 and 3

The fixed analysis of number of backward switches between emoji 1, 2 and 3 did not reveal any significant main effect or a significant interaction (see Table 11). The random effect of participant number showed a variance of .00 (*SD* = .00, indicating between-subject variability, and a residual variance of .24 (*SD* = .49, indicating within-subject variability).

Table 11. Fixed effects of the LMM predicting the number of backward switches between emoji 1, 2 and 3.

Predictor	**β**	**SD**	**t**	**p**
Intercept	.26	.23	1.13	.26
Sequence nature (unordered vs. ordered)	.05	.08	.62	.54
Order (scrambled vs. non-scrambled)	.02	.04	.40	.69
Sequence nature x Order	-.01	.06	-.22	.83
Categorical similarity score	.09	.12	.75	.45
Colour similarity score	-.09	.09	-.98	.33
Average ambiguity score	.03	.09	.34	.73

Predictor	β	SD	t	p
VLFI score	.00	.00	.95	.34
ELF2 score	-.00	.02	-.20	.84

df.resid = 1,1064

Significance levels: ***p < .001, **p < .01, *p < .05

Number of switches from emoji 1 to 2

The fixed analysis of number of switches from emoji 1 to 2 did not reveal any significant main effect or a significant interaction (see Table 12). The random effect of participant number showed a variance of .00 (*SD* = .00, indicating between-subject variability, and a residual variance of .18 (*SD* = .43, indicating within-subject variability).

Table 12. Fixed effects of the LMM predicting the number of switches from emoji 1 to 2.

Predictor	β	SD	t	p
Intercept	.34	.20	1.66	.10
Sequence nature (unordered vs. ordered)	-.04	.07	-.65	.52
Order (scrambled vs. non-scrambled)	-.05	.04	-1.40	.16
Sequence nature * Order	.04	.05	.77	.44
Categorical similarity score	.12	.10	1.01	.31
Colour similarity score	.02	.09	.28	.78
Average ambiguity score	-.07	.08	-.95	.34
VLFI score	-.00	.00	-.16	.87
ELF2 score	-.00	.02	-.14	.89

df.resid = 1,1064

Significance levels: ***p < .001, **p < .01, *p < .05

Number of switches from emoji 2 to 3

The fixed analysis of number of switches from emoji 2 to 3 did not reveal any significant main effect or a significant interaction (see Table 13). The random effect of participant number showed a variance of .00 (*SD* = .02, indicating between-subject variability, and a residual variance of .16 (*SD* = .40, indicating within-subject variability).

Table 13. Fixed effects of the LMM predicting the number of switches from emoji 2 to 3.

Predictor	β	SD	t	p
Intercept	.25	.20	1.28	.20
Sequence nature (unordered vs. ordered)	.03	.06	.47	.64

Predictor	β	SD	t	p
Order (scrambled vs. non-scrambled)	-.03	.04	-.88	.38
Sequence nature * Order	.03	.05	.55	.58
Categorical similarity score	.09	.10	.89	.37
Colour similarity score	-.05	.08	-.59	.55
Average ambiguity score	-.00	.07	-.04	.97
VLFI score	-.00	.00	-.63	.53
ELF2 score	-.02	.02	-.81	.42

df.resid = 1,1064

Significance levels: ***p < .001, **p < .01, *p < .05

Number of switches from emoji 2 to 1

The fixed analysis of number of switches from emoji 2 to 1 did not reveal any significant main effect or a significant interaction (see Table 14). The random effect of participant number showed a variance of .00 (*SD* = .00, indicating between-subject variability, and a residual variance of .15 (*SD* = .38, indicating within-subject variability).

Table 14. Fixed effects of the LMM predicting the number of switches from emoji 2 to 1.

Predictor	β	SD	t	p
Intercept	.17	.18	.95	.34
Sequence nature (unordered vs. ordered)	.02	.06	.38	.71
Order (scrambled vs. non-scrambled)	-.02	.03	-.55	.59
Sequence nature * Order	-.02	.05	-.40	.69
Categorical similarity score	.06	.09	.65	.52
Colour similarity score	-.04	.07	-.56	.57
Average ambiguity score	-.04	.07	-.50	.62
VLFI score	.00	.00	.79	.43
ELF2 score	.00	.02	.20	.84

df.resid = 1,1064

Significance levels: ***p < .001, **p < .01, *p < .05

Number of switches from emoji 3 to 1

The fixed analysis of number of switches from emoji 3 to 1 did reveal a marginally significant main effect of average ambiguity score (see Table 15), indicating that participants switched more from emoji 3 to 1 as average ambiguity ratings went up (see Figure 21).

The random effect of participant number showed a variance of .00 (*SD* = .01, indicating between-subject variability, and a residual variance of .04 (*SD* = .21, indicating within-subject variability).

Table 15. Fixed effects of the LMM predicting the number of switches from emoji 3 to 1.

Predictor	**β**	**SD**	**t**	**p**
Intercept	-.13	.10	-1.22	.22
Sequence nature (unordered vs. ordered)	-.00	.03	-.14	.89
Order (scrambled vs. non-scrambled)	.02	.02	1.33	.18
Sequence nature * Order	.01	.03	.25	.81
Categorical similarity score	.04	.05	.86	.39
Colour similarity score	-.01	.04	-.35	.72
Average ambiguity score	.07	.04	1.87	.06
VLFI score	.00	.00	.81	.42
ELF2 score	-.01	.01	-.70	.49

df.resid = 1,1064

Significance levels: ***p < .001, **p < .01, *p < .05

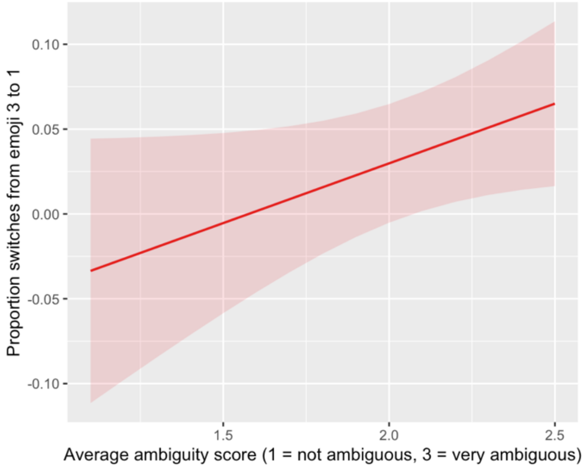

Figure 21. Proportion of switches from emoji 3 to 1, predicted for average ambiguity scores (1 = not ambiguous, 3 = very ambiguous), displayed with 95% confidence intervals.

Number of switches from emoji 3 to 2

The fixed analysis of number of switches from emoji 3 to 2 did not reveal any significant main effect or a significant interaction (see Table 16). The random effect of participant number showed a variance of .00 (*SD* = .01, indicating between-subject variability, and a residual variance of .15 (*SD* = .39, indicating within-subject variability).

Table 16. Fixed effects of the LMM predicting the number of switches from emoji 3 to 2.

Predictor	**β**	**SD**	**t**	**p**
Intercept	.23	.19	1.22	.22
Sequence nature (unordered vs. ordered)	.04	.06	.59	.56
Order (scrambled vs. non-scrambled)	.01	.03	.33	.74
Sequence nature * Order	.00	.05	.01	.99
Categorical similarity score	-.03	.09	-.26	.79
Colour similarity score	-.04	.07	-.56	.58
Average ambiguity score	-.01	.07	-.13	.89
VLFI score	.00	.00	.05	.96
ELF2 score	.00	.02	.00	1.00

df.resid = 1,1064

Significance levels: ***p < .001, **p < .01, *p < .05

Appendix III – First and third model predicting participants' judgements

Model 1

The fixed effects analysis of the first model predicting judgements revealed one main effect of sequence nature. As depicted in Figure 14, participants, were more likely to construct continuity for ordered sequences compared to unordered sequences.

Table 17. Fixed effects of the first LMM predicting judgements.

Predictor	**β**	**SD**	**t**	**p**
Intercept	1.91	.05	38.33	<.001***
Sequence nature (unordered vs. ordered)	.23	.05	4.75	<.001***
Order (scrambled vs. non-scrambled)	.01	.05	.24	.81
Sequence nature * Order	-.01	.07	-.20	.85

df.resid = 1,2394

Significance levels: ***p < .001, **p < .01, *p < .05

The random effect of participant number showed a variance of .05 (*SD* = .23, indicating between-subject variability, and a residual variance of .68 (*SD* = .82, indicating within-subject variability).

Model 3

The fixed effects analysis of the third model predicting judgements revealed four main effects of sequence nature, categorical similarity score, colour similarity score, and average ambiguity score (see Table 18). As depicted in Figure 14, participants, were more likely to construct continuity for ordered sequences compared to unordered sequences. Moreover, participants were more likely to construct continuity as categorical and colour similarity scores went up. In contrast, they were less likely to construct continuity as average ambiguity ratings went up.

Table 18. Fixed effects of the third LMM predicting judgements.

Predictor	**β**	**SD**	**t**	**p**
Intercept	2.38	.30	7.91	<.001***
Sequence nature (unordered vs. ordered)	.46	.05	9.88	<.001***
Order (scrambled vs. non-scrambled)	.01	.04	.14	.89
Sequence nature * Order	-.01	.06	-.22	.83

Predictor	**β**	**SD**	**t**	**p**
Categorical similarity score	-.43	.06	-6.83	<.001***
Colour similarity score	-1.01	.06	-15.75	<.001***
Average ambiguity score	.11	.05	2.38	.02*
VLFI score	-.00	.00	-.01	.99
ELF2 score	.04	.06	.74	.46

df.resid = 1,2389

Significance levels: ***p < .001, **p < .01, *p < .05

The random effect of participant number showed a variance of .06 (*SD* = .23, indicating between-subject variability, and a residual variance of .57 (*SD* = .75, indicating within-subject variability).