Language and Cognition (2022), 14: 4, 596–621
doi:10.1017/langcog.2022.19

ARTICLE

Picture perfect peaks: comprehension of inferential techniques in visual narratives

Bien Klomberg* and Neil Cohn

Department of Communication and Cognition, Tilburg School of Humanities and Digital Sciences, Tilburg University, Tilburg, The Netherlands
*Corresponding author. Email: s.a.m.klomberg@tilburguniversity.edu

(Received 31 May 2021; Revised 27 July 2022; Accepted 29 July 2022)

Abstract

The ability to reconstruct a missing event to create a coherent interpretation – bridging inference – is central to understanding both real-world events and visual narratives like comics. Most previous work on visual narrative inferencing has focused on fully omitted events, yet few have compared inference generation when climactic events become replaced with a panel employing numerous inferential techniques (e.g., action stars or onomatopoeia). These techniques implicitly express the unseen event while balancing several underlying features that describe their informativeness. Here, we examine whether processing and inference resolution differ across inferential techniques in two self-paced reading experiments. Experiment 1 directly compared five distinct types, and Experiment 2 explored the effect of combining techniques. In both experiments, differences in processing arise both between inferential techniques themselves, and at subsequent panels allowing the bridging inference to be resolved. Analysis of inferential features suggested that the explicitness of the inferential technique led to greater demand in processing, which later facilitated inference generation and comprehensibility. The findings reinforce the necessity of discussing the diversity of narrative patterns motivating bridging inferences within visual narratives.

Keywords: visual language; visual narrative; inference; metaphor

1. Introduction

Comprehending events is not always as straightforward as it may initially seem. Events consist of a string of actions, and observers may not witness each component part. Yet it is still possible to understand the full event, as one can often 'fill in' the missing information to make sense of it (Kosie & Baldwin, 2019). This retroactive construction of an unseen event is called a bridging inference (Hutson et al., 2018; Magliano et al., 2017; St. George et al., 1997). Studies of real-life events show that observers employ

bridging inferences quickly and that seeing only the buildup and the aftermath of an event is already sufficient to infer the main action (Strickland & Keil, 2011).

Such 'gap filling' is essential for understanding real-life events, but also for comprehending visual narratives like comics and picture stories (Cohn, 2019; Hutson et al., 2018; Magliano et al., 2016, 2017). Not only can visual narratives omit events to create bridging inferences (Hutson et al., 2018; Magliano et al., 2016, 2017), but the actual event may also be replaced by a panel that omits or implies the unseen action with a conventionalized inference-demanding technique (Cohn & Kutas, 2015; Cohn & Wittenberg, 2015). Consider Fig. 1a, which illustrates a simple sequential narrative,

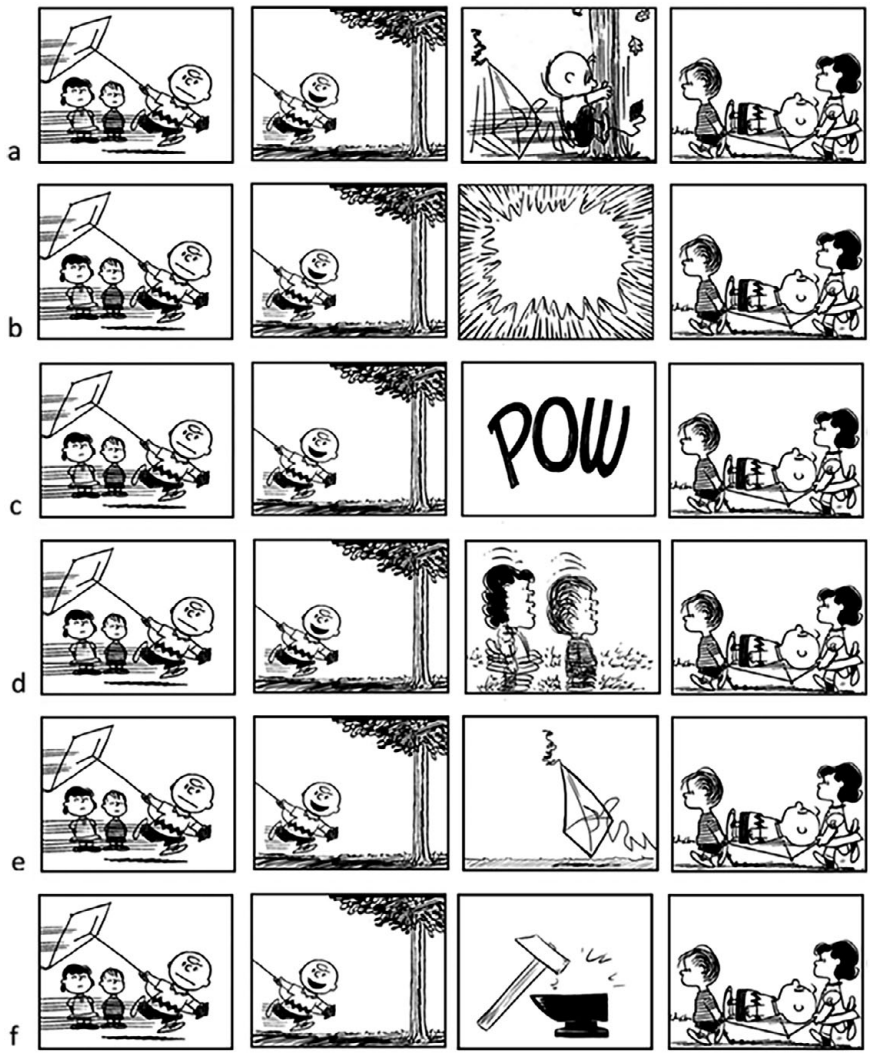

Fig. 1. Visual narrative sequences with (a) an explicit event, (b) an action star, (c) an onomatopoeia, (d) an echoic onlooker, (e) metonymic selective framing, and (f) a metaphor. Images are slightly adapted from *Peanuts* comics; *Peanuts* is © Peanuts Worldwide LLC.

https://doi.org/10.1017/langcog.2022.19 Published online by Cambridge University Press

and Fig. 1b, which shows a common inferential technique to substitute for the main action. In Fig. 1b, an 'action star' replaces the main event with a star-like symbol that indicates that something impactful happens, but does not provide cues for specific events. Even though there is still a panel in this position that cues that some event happens, readers need to generate a bridging inference to understand the full sequence.

Bridging inferences have received considerable focus in theories of comics (Cohn, 2019; Gavaler & Beavers, 2018; McCloud, 1993), where, unlike real life, an author can craft the presentation of an event omitted from a sequence. Several patterns can do this, each of which functions similarly in the narrative structure of a comic. Visual narrative grammar argues that panels play particular categorical roles, which may group together as a coherent constituents of a sequence (see Cohn, 2013b, 2020b). Consider the strips in Fig. 1. The Establisher (Panel 1) introduces the setting and the characters, before the Initial (Panel 2) sets up the climax of the scene, which occurs in the Peak (Panel 3). The aftermath or resolution of the event appears in the Release (Panel 4). This canonical order is stored in memory as a narrative schema (Cohn, 2014a, 2014b, 2020b). Studies have shown that readers recognize and dislike when a Peak has been omitted (Cohn, 2014b), highlighting their importance. Significant inferencing typically occurs when the contents of a Peak are missing, as this panel contains the primary events of the sequence contextually set up by the Initial panel and resolved by the Release panel (Cohn, 2019). What is interesting then about Fig. 1b is that the action star omits the primary events, yet it still fulfills the role of a Peak in the sequence. Thus, although an inference is required, the narrative structure remains intact.

Action stars are only one of many conventionalized inferential techniques that can replace Peaks and thus sponsor bridging inferences (Cohn, 2019). Along with action stars, this study focuses on four of these techniques, depicted in Fig. 1b–f. The Peak of Fig. 1c depicts an onomatopoeia, which is a sound effect evoked by the actual event, here a collision. In Fig. 1d, an 'echoic onlooker' depicts another character (or characters) viewing the event and re-enacting (part of) that event. Here, the onlookers, who watch the off-panel event, reproduce the vibrations that the protagonist experiences when slamming into the tree, despite not being a part of the event themselves. Fig. 1e portrays only a kite, a story element related to the Peak (since the kite falls down) and part of the scene (since the protagonist was holding it) but not showing the main action. This is an example of metonymic selective framing – something that conveys meaning (here, the event) by showing a related thing (here, the kite). Lastly, Fig. 1f illustrates a metaphor, which reflects the event with an abstract depiction of a comparable event. Here, a hammer beating an anvil is metaphoric for a collision, a (hard) hit between two elements, with the added similarity that both the anvil and the tree are, in general, immovable objects.

Although these inferential techniques all function structurally as Peaks in the narrative structure, they vary in how they imply undepicted content. Cohn (2019) posited that various features can describe the informativeness of each technique, as in Table 1. These features could be just descriptive theoretical constructs, or they may function psychologically in a 'preference rule' system described in approaches to the cognition of music or in theories of Gestalt psychology (Jackendoff & Lerdahl, 2006; Lerdahl et al., 1985). Preference rules are a characterization of principles involved in a dynamic system where features may exert various levels of strength in a context, thereby competing with each other to lead to an interpretation of a stimuli.

The features of inference techniques involve characteristics that cut across these structural patterns. The feature [blend] in Table 1 refers to establishing relations

https://doi.org/10.1017/langcog.2022.19 Published online by Cambridge University Press

Table 1. Overview of the distribution of features across inferential techniques

Inferential technique	Features						
	Blend	Framing	Explicit	Text	Substitutive	Arousal	Affix
Action star					X	X	X
Onomatopoeia			X	X	X	X	X
Echoic onlooker		X	X		X	X	
Metonymic selective framing	X	X	X			X	
Metaphor	X				X	X	

Note. This table demonstrates the presence of features across inferential techniques, with X indicating that the technique includes that feature.

between mental spaces (Fauconnier & Turner, 2002; Lakoff & Johnson, 1980) such as mapping across conceptual domains (metaphors) or within domains (metonymies). [Framing] refers to windowing part of the actual scene, as with metonymic selective framing and onlookers, which show parts of the scene other than the main events. [Explicit] techniques depict or describe aspects of the actual event. For example, echoic onlookers, onomatopoeias, and metonymic panels all directly relate via mimicking movement, evoking sound, or showing a part of the implied event. [Text] indicates the presence of text. [Substitutive] techniques show something other than the event, so a replacement of the original scene. All selected inferential technique are events and therefore have high [arousal], as opposed to states, which would have low [arousal]. Lastly, [affix] indicates if morphological affixes are used, which are entities that typically connect to other elements, such as the action star which canonically connects to the 'stem' that generates that 'flash' (i.e., connects to the source of the impact; Cohn, 2013a).

When a reader progresses through a (visual) discourse, they incorporate the information into a situation model: a mental representation of the ongoing narrative, including the events, actions, and agents involved (Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). This type of mental model construction has been studied extensively in research on verbal discourse (Graesser et al., 1994; Kuperberg et al., 2010; Yang et al., 2007) and in the comprehension of real-life events (Papenmeier et al., 2019; Zacks et al., 2007). Situation models update with changes to the described or depicted events. Bridging inferences reflect such an update to the situation model when meaning is missing and needs to be filled in.

Construction of a situation model involves several operations. The scene perception and event comprehension theory (SPECT) by Loschky et al. (2020)) describes that visual narrative understanding involves front-end processing where readers extract information based on the content of the panel, such as the entities and events involved. These link to back-end processes where extracted information activates representations encoded in semantic memory, which feed into the construction of an event model. The parallel-interfacing narrative semantics (PINS) model by Cohn (2020b) adds that incoming information from each image prompts predictions about upcoming information. For example, a reader may assume that a character from the first panel will persist in the following panels as well. Sequences that adhere to expectations facilitate processing (Coderre et al., 2020; Cohn, 2020b). Both SPECT and PINS establish that each consecutive panel is integrated in the existing situation model via updating processes. The more

https://doi.org/10.1017/langcog.2022.19 Published online by Cambridge University Press

incongruent the incoming stimuli is, the greater the update that is required (Huff et al., 2014; Magliano & Zacks, 2011).

Bridging inferences invite such updating processes through the need to construct a coherent event understanding in the situation model (Graesser et al., 1994). Such updating is prompted not only by dropping out panels with crucial event information, like a Peak (Hutson et al., 2018; Magliano et al., 2016, 2017), but also when encountering a Peak panel without explicit information (as in Fig. 1b–f). Indeed, readers slow down when reading panels following action stars or blank panels which replace climactic events (Cohn & Wittenberg, 2015). Overall, longer self-paced reading times have been taken as evidence for inference generation mechanisms in both visual and verbal narratives, and some work has implicated overlapping working memory processes operating across both modalities (Magliano et al., 2016).

The SPECT model keeps to semantic processing, but the PINS model includes narrative structure as an equally important component of comprehending visual sequences. The visual cues of a panel and its position in the narrative help to establish its narrative category. For example, preparatory actions likely correspond to an Initial panel, which entails expectations for the upcoming stimuli to follow the canonical order as a Peak. When the subsequent panel is then observed, backward revision processes will confirm or revise the interpretation of the current structure. Thus, even though the inferential techniques in Fig. 1b–f should sponsor inferencing, they should all maintain the narrative structure as Peaks.

Nevertheless, it remains unknown whether these inferential techniques differ in how they encourage readers to infer unseen information. While the processing of some inferential techniques has been explored, little research has compared their comprehension. Therefore, this study examines to what extent processing differs across conventionalized inferential techniques. One possibility is that, while differences may persist between techniques, they may be motivated by the features (as in Table 1) used to describe their abstract similarities and differences (Cohn, 2019). When such features were theorized, it was unclear whether they served a purely descriptive, theoretical function or whether they could characterize psychological constructs involved in processing. Thus, this study also explores how processing may vary on the basis of underlying features that cut across inferential techniques.

To examine these issues, two experiments measured participants' self-paced reading of comic strips. Experiment 1 first compares the five outlined inferential techniques (Fig. 1), whereas Experiment 2 then examines the effect of combining inferential techniques.

2. Experiment 1: comparing inferential techniques

While various studies have investigated the processing of visual narratives when events are fully missing (Hutson et al., 2018; Magliano et al., 2016, 2017), comparison of inferential techniques remains limited. Cohn and Wittenberg (2015) compared the self-paced reading of sequences with explicit events with those with action stars, blank panels, and anomalous Peaks. Even though blank panels contained no visual information at all, action star panels were viewed faster than blank panels. This shorter viewing time was taken as evidence that action stars played narrative roles as Peaks, unlike blank panels. However, panels after action stars or blank panels both evoked equally longer viewing times, suggesting that the lack of

https://doi.org/10.1017/langcog.2022.19 Published online by Cambridge University Press

explicit cues in both action stars and blank panels necessitates updating of the situation model. Recent work has further observed little difference in the brain responses between panels following action stars and 'noise' panels with scrambled lines (Cohn, 2021).

Such semantic cues may potentially mediate comprehension through the explicitness of inferential techniques. Cohn and Kutas (2015) compared onlookers with varying amounts of cues about the off-panel event they were watching. Neural responses indicative of updating processes were larger for onlookers with more cues about the event, such as matching facial expressions and an exclamation mark above their head, than those without those cues. Explicit depictions of events had even greater responses. Thus, the more information that was depicted in the critical panel, the more updating that was required. At the subsequent panel, onlooker panels evoked brain responses different from explicit depictions in a way that suggested the possibility of working memory processes involved in inference.

Some work has also examined substitutions of text for events. Huff et al. (2020) compared textual descriptions of events to visually explicit or omitted events within a visual narrative sequence. The text required longer viewing times than the visuals, suggesting that switching modalities may require more effort than unimodal sequences. In addition, when sound effects are shown instead of an event, processing is impacted by the type of word and its congruity. Descriptive words (e.g., 'punch') substituted for Peaks evoke brain responses suggestive of being more unexpected than onomatopoeic words (e.g., 'pow') and words of either type that were incongruent to the sequence context (e.g., 'kiss' or 'smooch' substituting a punch) were more costly to semantic processing (Manfredi et al., 2017).

No studies have yet looked at the processing of metaphors or metonymic selective framing techniques in visual narrative sequences. Visual metaphors broadly have only recently begun receiving empirical attention, and many studies focus on a comparison to verbal metaphors (Ojha et al., 2019), rather than to other visual techniques. Still, metaphoric images from advertisements require more processing costs than literal advertisement images (Ortiz et al., 2017). Studies of language further suggest that both metaphor and metonymies could lead to reading costs. For metaphors, familiarity is often considered decisive. Giora (2003) posits that salient (the most consolidated) interpretations of words are always activated most strongly, whether this is the literal or metaphoric interpretation. Another factor is context; metaphors in context can be understood as quickly as literal phrases when of good quality (Glucksberg, 2003; Glucksberg & McGlone, 2001). Quality, or how well the metaphor expresses the target, is especially important for novel metaphors (Brisard et al., 2001). Moreover, de Vries et al. (2018) observed that metaphors with clear metaphoric markers (e.g., 'as if') were read slower than metaphors without markers and than non-metaphorical expressions.

Similar factors seem involved in processing metonymic devices, as both familiarity and context facilitate interpretation (Frisson & Pickering, 1999, 2007). Despite these similarities, metonymies appear less complex than metaphors, and are comprehended easier (Rundblad & Annaz, 2010). This difference prompts the prediction that for visual narratives too, metonymic panels may remain easier to interpret than metaphoric ones. Furthermore, unlike metaphors, metonymic selective framing also varies in terms of framing, since they highlight specific

https://doi.org/10.1017/langcog.2022.19 Published online by Cambridge University Press

elements of the scene. Similarly, Foulsham and Cohn (2020) created panels that zoomed in only on the parts of an image that had been fixated on by a previous group of observers. Such fixation zoom panels were easier to understand if they included informative cues of a scene rather than uninformative cues. These findings together suggest that metonymic selective framing panels may be comprehended faster than metaphor panels.

Altogether, only a few inferential techniques have been explored in the processing of visual narratives, and rarely compared. Here, we thus ask to what extent the processing of inferential techniques differ from each other by comparing action stars, onomatopoeia, echoic onlookers, metaphoric selective framing, and metaphoric panels (Fig. 1). We predict that both front-end and back-end factors will influence participants' self-paced reading of these inferential techniques.

First, related to front-end processing, we predict that panels with fewer visual features will be viewed faster than more complex representations. SPECT predicts that information extraction processes are based on eye fixations, meaning that less visual content to be extracted should result in fewer fixations and thus faster viewing times. Thus, action stars and onomatopoeias will be viewed faster than other conventionalized inferential techniques (see Cohn & Wittenberg, 2015). In addition, subsequent back-end processing would predict that explicitness should factor into processing (Cohn, 2019; Cohn & Kutas, 2015), because more explicit cues for events should better assist constructing a situation model. Thus, the more explicit echoic onlookers and metonymic selective framing panels may be viewed slower than inexplicit metaphors.

Nevertheless, inferential processing should be evident at the panel after the Peak, where viewing times should be longer for the panel following an inferential peak than the one following an explicitly depicted event (Cohn & Wittenberg, 2015; Magliano et al., 2016, 2017). Given that prior work has found similar viewing times across inferential techniques (Cohn & Wittenberg, 2015), one possibility is that all the panels following inferential peaks may sponsor similar viewing times. This would imply that the bridging inference needed to reconcile missing information would not differ, no matter how that information is implied in different techniques. Another possibility might follow that if explicitness provides greater access to event representations at the inferential Peak itself, processing at the subsequent image should become easier for techniques that are more explicit than those that are less explicit.

Finally, by directly comparing inferential techniques, we sought to examine whether their proposed underlying features indeed function as psychological constructs. We predicted that techniques with an [explicit] feature would lead to slower viewing times at the critical panel, based on the cost of accessing event structures by explicit cues, but which would then speed up viewing times for the subsequent panel (Cohn, 2019; Cohn & Kutas, 2015; Cohn & Paczynski, 2013). Due to the costs for switching modalities (Huff et al., 2020), [text] features may slow down readers at both the critical and subsequent panels. Then, based on SPECT, [framing] may imply a reduction of visual features, which could lead to reduced fixations in front-end processing, ultimately leading to faster viewing times at the critical panel. Lastly, metaphor and metonymic panels share a [blend] feature, which would be predicted to slow down readers both at the critical and subsequent panels, reflecting the costs of mappings between mental spaces hinted at in studies of language.

https://doi.org/10.1017/langcog.2022.19 Published online by Cambridge University Press

3. Methods

3.1. Stimuli

We selected 30 four-panel *Peanuts* sequences with no words from an existing stimulus set (Cohn, 2019; Cohn & Kutas, 2015). For each strip, based on the events of the original Peak panel, five additional panels were designed for each of the inferential techniques (action star, onomatopoeia, echoic onlooker, metonymic selective framing, and metaphor). These panels were created by editing the original panels, using other panels in the database, or drawing new panels that matched the style of *Peanuts*. The total number of stimuli was thus 180 strips (30 strips × 6 sequence types, as shown in Fig. 1). These sequences were counterbalanced into six lists using a Latin Square Design, such that each participant viewed 30 strips in total and each strip appeared only once with five examples of each sequence type, but all strips in all sequence types were viewed across participants.

For all 30 strips, a cloze probability score (see Coderre et al., 2020) and inference assessment score (see Cohn, 2021) were measured in rating studies. For all participants, we measured their Visual Language Fluency Index (VLFI), a metric developed to assess comic reading expertise (Cohn, 2020a), with scores above 20 indicating a high comic fluency (i.e., the reader is well versed in reading comics), scores below 7 a low comic fluency, and an ideal average of around 12. For cloze probability, 47 participants (27 female; mean age: 21.8, range: 18–32; mean VLFI: 8, range: 1.5–17.6) viewed 30 comics with only the first two panels, and they were asked to describe what happened next. The average cloze probability score was 0.38 (range: 0–0.87), meaning that on average 38% of participants predicted the correct event. A consensus score was also included, which measured how often participants agreed with another, regardless of their accuracy to the actual subsequent events. On average, the cloze consensus score was 0.48 (range: 0.17–0.87). For the inference assessment score, 49 participants (39 female; mean age: 21.3, range: 18–35; mean VLFI: 11.3, range: 1.5–38.5) viewed the same 30 strips, with the Peak omitted (always the third panel). They were then asked to describe what happened in the missing panel. The inference assessment score ranged from 0.04 to 0.98, averaging 0.60. Here too, a consensus score was included, with a range between 0.29 and 0.98, averaging 0.67.

The inference assessment showed that three strips were too hard to infer, which were therefore removed from analyses. The cloze probability scores then ranged from 0.02 to 0.87, with an average of 0.41 (consensus range: 0.23–0.87, mean: 0.50). The inference assessment scores then ranged from 0.35 to 0.98, averaging 0.68 (identical for the consensus scores). So, even though the majority of the strips had low cloze, with the subsequent Release, most events remained inferable.

3.2. Participants

A total of 117 participants were recruited via social media and the participant pool available through Tilburg University, the Netherlands. The mean age was 29.33 years ($SD = 12.05$, range: 17–64, 51 male, 61 female, 5 other). Although statistical power was not calculated a priori, a post hoc power analysis in G*Power indicated that with 12 conditions across 117 participants, to achieve a medium effect size of 0.25, it required F-values of above 1.80 for our within-subjects design. Likewise, with six conditions, we required F-values of above 2.23. These values were met, suggesting

https://doi.org/10.1017/langcog.2022.19 Published online by Cambridge University Press

that our sample size yielded sufficient power. All participants gave their informed written consent according to protocols approved by the Tilburg University School of Humanities and Digital Sciences Research Ethics and Data Management Committee. All participants completed the VLFI (Cohn, 2020a). The mean VLFI score for this sample was average, at 13.83 ($SD = 9.49$, range: 1.5–42.5).

3.3. Procedure

Participants accessed the experiment via an online link. First, they viewed an introductory text with instructions and answered the VLFI questions. During the following experimental part, the stimuli were presented in a self-paced viewing set-up via Qualtrics, using the lab.js JavaScript plugin (Henninger et al., 2022). Before each sequence, a screen indicated which trial the participants were at, for example, trial 1 out of 30. Each trial consisted of a four-panel sequence which appeared one panel at a time at the participants' own pace by pressing the spacebar. After each sequence, participants rated its coherence by pressing the keyboard (1 = hard to understand to 7 = easy to understand). After all 30 sequences, participants were debriefed and thanked for their contribution. On that final page, they could also report if they had noticed anything unusual or had any additional comments.

3.4. Data analysis

Viewing times below 300 ms and above 8,000 ms were first excluded for being either too fast or too slow following Cohn and Wittenberg's (2015) approach, which based these cut-off values on previous self-paced reading results with the mean fastest and slowest viewing times well between these limits (Cohn, 2014b; Cohn & Paczynski, 2013). Outliers were calculated as 2.5 times the standard deviation above the mean, resulting in two participants' responses being excluded, for a final sample of 117 participants' responses. Viewing times and comprehensibility ratings were averaged across items for each participant.

First, we analyzed viewing times in a 2 (Position: critical panel and critical panel +1) × 6 (Sequence Type: original event panel, action star, onomatopoeia, echoic onlooker, metonymic selective framing, and metaphor) factorial analysis of variance (ANOVA), to explore whether the position and the type of Peak panel affect viewing times.

For more in-depth analysis, we examined viewing times at the critical Peak panel and at the subsequent panel (critical panel +1) separately, and the comprehensibility ratings for the whole sequence. The data were analyzed using repeated-measures ANOVAs, with Peak type as independent variable (six levels: original event panel, action star, onomatopoeia, echoic onlooker, metonymic selective framing, and metaphor). A Greenhouse–Geisser correction was used at the Peak panel and for the ratings to account for the violation of the Mauchly's test of sphericity. We report the corrected p-values and corrected degrees of freedom. Post hoc analyses used a Bonferroni correction for multiple comparisons. Next, regressions were used to examine whether the different features functioned as predictors for viewing times at critical panels and for comprehensibility ratings. VLFI scores were correlated with the difference between the viewing times of inferential sequence types and those of the original sequence, to see an influence of comic

https://doi.org/10.1017/langcog.2022.19 Published online by Cambridge University Press

expertise. The data and analyses have been made available in an online data repository (https://doi.org/10.34894/DTBW7M).

4. Results

4.1. Viewing times

We analyzed viewing times in a 2 (Position: critical panel and critical panel +1) × 6 (Sequence Type: original event panel, action star, onomatopoeia, echoic onlooker, metonymic selective framing, and metaphor) factorial ANOVA. There was a main effect of position, $F(1, 1{,}392) = 281.62$, $p < 0.001$, $\eta\text{partial}_2 = 0.16$, with slower viewing times for panels at the critical panel +1 position (M = 1,902.13, SD = 862.21) opposed to panels at the critical Peak panel position (M = 1,255.58, SD = 597.11). There was also a main effect of sequence type, $F(5, 1{,}392) = 9.04$, $p < 0.001$, $\eta\text{partial}_2 = 0.03$. Post hoc analyses outlined that the onomatopoeia (M = 1,403.90, SD = 731.54) was viewed faster than metonymic selective framing (M = 1,600.67, SD = 762.06), echoic onlookers (M = 1,672.72, SD = 857.23), and metaphors (M = 1,793.71, SD = 880.66). Likewise, the original event panel (M = 1,465.97, SD = 664.85) was viewed faster than echoic onlookers and metaphor, and finally, the action star (M = 1,536.18, SD = 877.30) and metonymic selective framing panel were read faster than the metaphor. Last, there was also an interaction effect, $F(5, 1{,}392) = 8.54$, $p < 0.001$, $\eta\text{partial}_2 = 0.02$. To explore our findings here, we consider each position in their own analysis.

Fig. 2 shows the viewing times for the critical panel and critical panel +1 for all six sequence types at both panel positions. Viewing times were then analyzed at the third

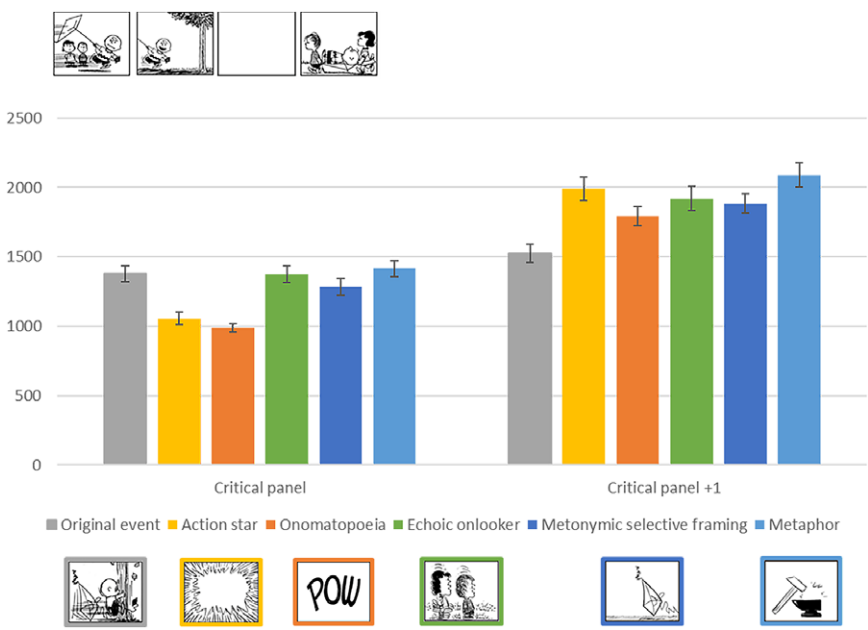

Fig. 2. Overview of viewing times at the critical Peak panel and subsequent panel for all six sequence types; the error bars represent standard errors.

https://doi.org/10.1017/langcog.2022.19 Published online by Cambridge University Press

panel position, the critical Peak panel. There was a main effect of sequence type on the response times, $F(4.34, 503.45) = 35.94$, $p < 0.001$, $\eta\text{partial}_2 = 0.24$. Post hoc analyses showed that both action stars and onomatopoeias were viewed faster than echoic onlookers, metonymic selective framing, metaphors, and original event panels (all $ps < 0.001$). Metonymic selective framing panels were faster than metaphors ($p = 0.041$).

At the critical panel +1, the sequence type was shown to have a main effect on the response times, $F(4.63, 536.67) = 23.60$, $p < 0.001$, $\eta\text{partial}_2 = 0.17$. The panels following original event Peaks were viewed faster than those after all inferential techniques (all $ps < 0.001$). Panels after onomatopoeias were viewed faster than those after action stars ($p = 0.025$) and metaphors ($p < 0.001$). Moreover, panels following echoic onlookers and metonymic selective framing were both viewed faster than those after metaphors ($p = 0.037$ and $p < 0.001$, respectively).

4.2. Comprehensibility rating

The sequence types differed in their comprehensibility ratings, $F(4.29, 497.95) = 84.93$, $p < 0.001$, $\eta\text{partial}_2 = 0.42$. All conditions differed from each other (all $p < 0.001$) except for ratings between action stars and metonymic selective framing, and for ratings between echoic onlookers and metaphors (Fig. 3). Sequences with original event panels were rated most comprehensible (all $p < 0.001$), then sequences with onomatopoeias, which were rated higher than those with other inferential techniques (all $p < 0.021$). Last, sequences with action stars and metonymic selective framing were both rated more comprehensible than echoic onlookers and metaphors (all $p < 0.001$).

4.3. Features analysis

A multiple regression was used to investigate the predictive power of the underlying features for viewing times at the critical Peak panel, the critical panel +1, and for the

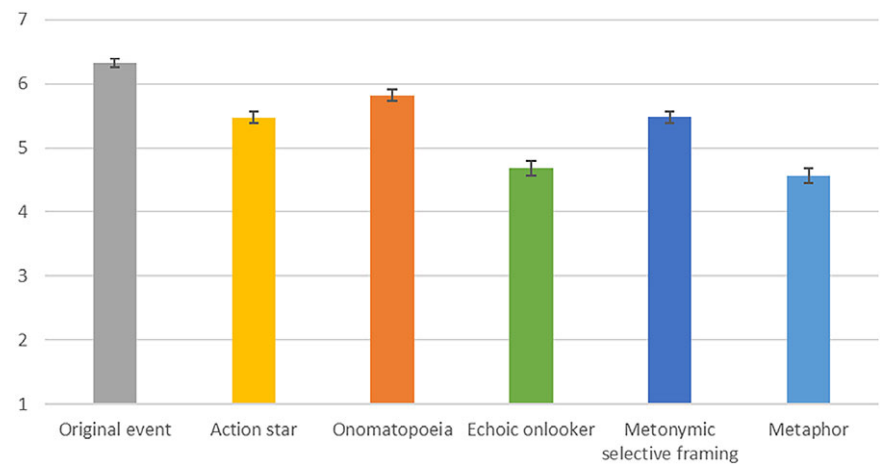

Fig. 3. Overview of the comprehensibility ratings for all six sequence types; the error bars represent standard error.

https://doi.org/10.1017/langcog.2022.19 Published online by Cambridge University Press

comprehensibility ratings. Due to collinearity, the features [arousal] and [substitutive] were excluded from analysis. As [text] was unique to onomatopoeias, it was left out as well. Likewise, [affix] only pertained to onomatopoeias and action stars, which are affixes by definition, and did not apply to any other sequence type. Therefore, the final analysis included only the features [blend], [framing], and [explicit]. These features were then also correlated against one another to test their relationship, and appeared as valid predictors with a shared variance of .25 at most. To further examine their relative influence, we also conducted general dominance and relative importance analyses; the complete output can be found in the online repository.

At the critical panel, the regression showed that [blend], [framing], and [explicit] explained 28.2% of the variance in the viewing times ($R^2 = 0.08$, $R^2_{\text{Adjusted}} = 0.06$, $F(3, 152) = 4.38$, $p = 0.005$). Table 2 reports the *t*-values and *p*-values of each feature, and Fig. 4 graphs the positivity or negativity of the standardized betas. The feature [blend] led to slower viewing times. At the critical panel +1, the features explained 37.5% of the variance in viewing times ($R^2 = 0.14$, $R^2_{\text{Adjusted}} = 0.12$, $F(3, 152) = 8.29$, $p < 0.001$). Here, [framing] predicted slower viewing times, and [explicit] led to faster viewing times. For the comprehensibility

Table 2. Overview of *t*-values and *p*-values for each feature per dependent variable

	Critical panel		Critical panel +1		Ratings	
Feature	*t*-value	*p*-value	*t*-value	*p*-value	*t*-value	*p*-value
Blend	2.43	0.016	0.51	0.613	−0.39	0.701
Framing	1.26	0.209	2.39	0.018	−5.54	<0.001
Explicit	0.37	0.709	−4.18	<0.001	5.92	<0.001

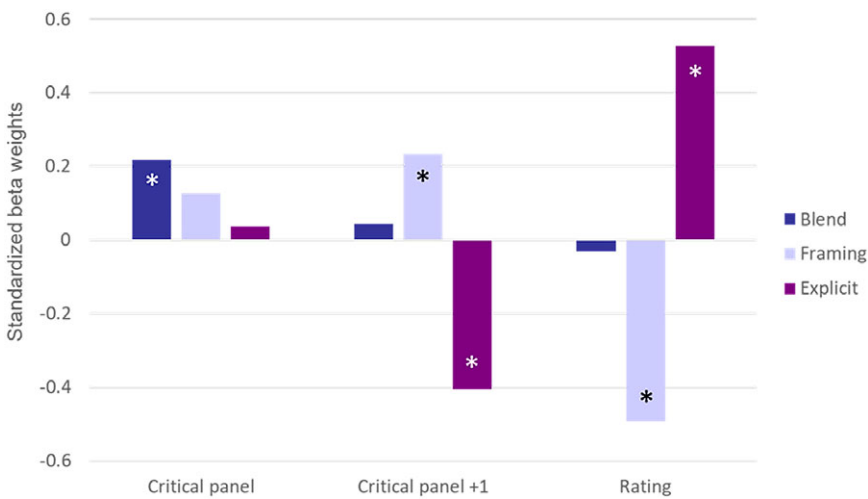

Fig. 4. Beta-weights from a regression examining the influence of different features on the viewing times and self-rated comprehension of the sequence.

https://doi.org/10.1017/langcog.2022.19 Published online by Cambridge University Press

ratings, the features explained 52.7% of the variance ($R^2 = 0.28$, $R^2_{\text{Adjusted}} = 0.26$, $F(3, 152) = 19.48$, $p < 0.001$). [Framing] predicted lower ratings, whereas [explicit] predicted higher ratings.

In addition to the inferential features, the cloze probability and the inference assessment scores were tested as predictors, but no significant relations emerged. In addition, we explored the relationship between inference assessment scores and comprehensibility ratings to assess whether better understanding of the sequence was related to the transparency of the missing event. Specifically, action stars were rated more comprehensible than echoic onlookers and metaphors, even though action stars remain the least explicit, giving more of an opportunity for readers to fill in the meaning. Thus, one could interpret action stars as merely signals to draw an inference, and comprehensibility ratings could inform how easy readers inferred those events. However, no relations emerged between these ratings and the inference assessment scores. An alternative possibility is that action stars allow readers to fill in information without additional conflict from explicitly presented information. The other suppletive Peaks may provide information that could conflict with a reader's internally generated interpretation. For example, the incoming information in metaphor sequences may conflict with the internally represented event information, which gives way to an inference (metaphor) to explain that relationship. Action stars being fairly inexplicit, however, may readily connect to internal representations.

Furthermore, we correlated inference assessment scores also with ratings across all sequences, which showed that higher comprehensibility ratings aligned with lower inference assessment scores ($p = 0.037$). Potentially, when the first two panels of a sequence did not elicit a clear internal interpretation, an implicit cue (i.e., the inferential technique) aided understanding. In contrast, for those sequence that were easy to infer the implicit cue may have conflicted with the internally generated interpretation, decreasing comprehensibility.

4.4. Visual language fluency

To examine the influence of comic reading expertise on the magnitude of response, we correlated VLFI scores with the difference between the viewing time or rating of the original event panel subtracted from that of each inferential technique. At the critical panel, the difference between original event panels and echoic onlookers was greater for more experienced comic readers than for less experienced readers, $r(115) = 0.20$, $p = 0.035$. As the mean difference score for echoic onlookers was a negative value, this suggests that echoic onlookers were read faster than the original event panels. At the critical panel +1, higher comic expertise indicated slower viewing times for panels after metaphoric than original Peaks, $r(115) = 0.22$, $p = 0.017$, as there, the mean difference score for metaphors was a positive value. For comprehensibility ratings, there were no significant correlations.

5. Discussion

This experiment compared the self-paced viewing times of inferential techniques in visual narratives. We found that all panels following an inferential technique were

https://doi.org/10.1017/langcog.2022.19 Published online by Cambridge University Press

viewed longer than the panel following original event panels, but with varying differences across sequence types. Inferential panels with generally fewer visual cues (action stars and onomatopoeia) were viewed faster than those with more details (echoic onlookers and metaphors); metonymic selective framing was left somewhat in the middle. Analysis of features suggested that faster processing at the subsequent panel aligns with higher comprehensibility ratings, whereas slower viewing times align with lower ratings. Overall, narrative techniques motivate inference generation in different ways, and features persisting across these techniques may influence their comprehension.

At the Peak panel, the differences in viewing times between inferential techniques did not necessarily indicate variance in inference generation. Rather, the most salient difference between techniques was the faster viewing times for action star and onomatopoeia panels. This could raise the question whether these were simply easier to process based on their input or whether readers instead did not process these as deeply, a distinction between intrinsic and germane cognitive load (Sweller, 2010). The consistently longer viewing times at the subsequent panel suggested that these viewing time differences did not pertain to deepness of processing, but rather to differences in surface-level structures. The faster viewing times made to action stars and onomatopoeias specifically seemed related to visual complexity of the panels themselves (Cohn, 2021; Cohn & Wittenberg, 2015). The results supported SPECT's (Loschky et al., 2020) predictions that less content should require fewer eye fixations, and thus result in faster viewing times. This may also explain the relatively fast processing of metonymic selective framing panels – which showed reduced visual content (e.g., parts of characters/objects) and/or motion lines – compared to more complex content (echoic onlookers and metaphors).

Nevertheless, our analysis of features showed that visual complexity alone did not predict the viewing time results. Here, the [blend] feature predicted slower viewing times of Peak panels, which fits with the slowly processed metaphors and to some extent with metonymic selective framing. Evidently, effects at the Peak do not necessarily carry through to subsequent panels or comprehensibility, as this feature appeared of no influence there. Seemingly, at this point in the sequence, viewing times may be influenced more by the quantity of visual cues than the nature of those cues.

Evidence of inferential processing appeared at the subsequent panel where the viewing times of panels following inferential techniques were all longer than following the original event panel (Cohn & Kutas, 2015; Cohn & Wittenberg, 2015; Huff et al., 2020). Here again, viewing times differed between inferential techniques. The fastest viewing times came to panels that followed explicit inferential techniques (onomatopoeia, echoic onlookers, and metonymic selective framing), suggesting that although this explicitness did not necessarily affect initial processing at the Peak panel, it benefited comprehension at the subsequent panel. Conversely, the relatively *inexplicit* action stars and metaphors had the slowest viewing times. Comparing the explicit onomatopoeia and inexplicit action star, both similar in their visual complexity, at this subsequent panel, further supports how readers process the information of these Peaks deeply enough and how those cues become relevant predominantly when an inference is prompted. Namely, the onomatopoeia expressed a distinct sound effect most appropriate for a particular event, which makes it more informative than the identical action star across sequences. This facilitated inference resolution to a larger degree, so

https://doi.org/10.1017/langcog.2022.19 Published online by Cambridge University Press

similarly simple content can still have varying consequences. Action stars provide little semantic information at all, thus requiring updating at the final panel (Cohn & Wittenberg, 2015), but panels after metaphors evinced the longest viewing times. Although metaphoric panels show explicit events, on the surface they are incongruous, not showing events that actually occur in the sequence. Since metaphors were not viewed slower than the congruous original event panels at the Peak position, it suggests that at this moment readers do not consider them fully anomalous (Cohn & Wittenberg, 2015). Rather, this non-sequitur information needs to be resolved with the sequence's events in order to remain congruent. Indeed, this may not have always been achieved, as sequences with metaphoric Peaks had the lowest comprehensibility ratings. The panels following metaphors also correlated with comic reading expertise, such that more fluent comic readers spent more time on them. Metaphors thus appear a relatively difficult inferential device, and perhaps especially odd as a Peak for those more familiar with comics.

Contrary to the Peak panel, the [blend] feature did not predict viewing times at the subsequent panel, but [framing] did predict slower viewing times despite not influencing the preceding panel. This may explain why the panels after echoic onlookers and metonymic selective framing (both with [framing] features) were slower than those after onomatopoeia, despite all being explicit. Both our standardized beta-weights and our additional analyses of dominance and importance (see online repository) suggest that features may exert various levels of strength in relationship to one another, like 'preference rules' in dynamic systems described in approaches to the cognition of music or in theories of Gestalt psychology (Jackendoff & Lerdahl, 2006; Lerdahl et al., 1985). As this is the first study to investigate these inferencing techniques in this way, future studies can better assess the relative influences and constraints on these features.

Explicitness also predicted the comprehensibility ratings to inferential techniques, which were all below ratings to the original sequences. Both onomatopoeias and metonymic selective framing were rated high. Only the explicit echoic onlookers scored low, which could be due to foregrounding more characters than other techniques, complicating the sequence. Moreover, in some sequences, echoic onlookers were introduced in the background of previous panels, as in Fig. 1d, but in others the onlooker(s) appeared for the first time in the Peak. Readers thus may have disliked encountering novel characters this late in the sequence, since unexpected entities require more mental model updating (Cohn & Kutas, 2015; Reid & Striano, 2008).

Among the inexplicit techniques, metaphors were also rated low, possibly due to the extended updating needed to interpret them, whereas action stars were rated high. Their ratings may have been motivated by their familiarity, as conventionalized units within comics (Cohn, 2021; Cohn & Wittenberg, 2015), arguably more so than metaphors. Moreover, action star sequences were viewed faster than metaphor sequences, and potentially comprehensibility ratings reflect the effort required for inference resolution. The more effort the reader had to put in, the less understandable they perceived it to be.

Finally, the analysis of features offered insight into the flow of information across viewing times at the Peak, the subsequent panel, and comprehensibility ratings. While most features appeared of no influence at the Peak, converse trends appear afterward: a feature predicting slower times at the subsequent panel also

https://doi.org/10.1017/langcog.2022.19 Published online by Cambridge University Press

predicted lower ratings, and faster times aligned with higher ratings. These results are broadly consistent with theories that greater access of information in one panel would lead to subsequent facilitation (or vice versa) across the course of the sequence (Cohn, 2020b; Loschky et al., 2020), overall affecting comprehensibility.

6. Experiment 2: comparing combinations of techniques

Experiment 1 showed differences in processing and comprehensibility between inferential techniques. Yet, features stretching across the conventionalized techniques interacted to exert relative strengths, which we aimed to explore further. One of the most prominent features affecting inferences and comprehensibility was explicitness, and this feature also underlies the onomatopoeia, an inferential technique readily combinable with other panels. Therefore, to assess additive or competing features, Experiment 2 compared combining onomatopoeia with action stars, echoic onlookers, metaphors, and original event panels (see Fig. 5), to their 'silent' versions (as shown in Fig. 1).

Adding onomatopoeias to panels makes them multimodal. Switching modalities to a fully textual panel may incur costs to recode information to fit the visually based mental model of the rest of the sequence (Huff et al., 2020). However, this effect appeared for phrasal descriptions of events rather than single word sound effects. Indeed, Manfredi et al. (2017) found that inferential panels with only descriptions of events ('Punch!') were regarded as more unusual than onomatopoeias ('Pow!'). Typical onomatopoetic (nondescriptive) sound effects are a common device in comics (Guynes, 2014; Pratha et al., 2016), and as such, could be easier to digest than a fully phrasal description. In addition, sound effects can take on image-like qualities, styled to fit the visuals in a way unlike other text elements (Pratha et al., 2016). Thus, while panels with onomatopoeia are multimodal, sound effects may not incur as much a cost of switching modalities as other text. Other work has demonstrated that processing costs interface with congruity (Manfredi et al., 2018), so as long as the multimodal combinations are congruous, they are processed easier.

Experiment 2 thus explores the processing and comprehension of combining onomatopoeia with other inferential techniques. Based on Experiment 1, multimodal versions would be viewed faster than unimodal versions due to the added effect of the [explicit] feature. However, as the additional visual complexity of multimodal versions compared to unimodal versions likely increases viewing times as well, we ultimately expect slower viewing times for panels including onomatopoeias. At the

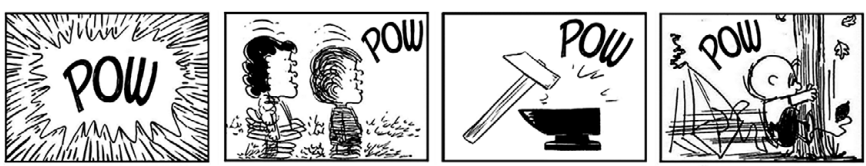

Fig. 5. Examples of the onomatopoeia combination panels for (a) action stars, (b) echoic onlookers, (c) metaphors, and (d) original event panels. The images are slightly adapted from *Peanuts* comics; *Peanuts* is © Peanuts Worldwide LLC.

https://doi.org/10.1017/langcog.2022.19 Published online by Cambridge University Press

subsequent panel, the explicitness of onomatopoeias should lead to shorter viewing times, further facilitating the bridging inferences, and thus leading to higher comprehensibility ratings.

7. Methods

7.1. Stimuli

Experiment 2 used the same 30 strips from Experiment 1 along with two additional strips to sufficiently counterbalance the number of strips per condition. The Peak panels were shown either as in Experiment 1 (Fig. 1) or combined with an onomatopoeia (Fig. 5). Consequently, no sequence type used only onomatopoeia panels, and metonymic selective framing panels were also excluded. This resulted in a 2 (Modality: unimodal and multimodal) × 4 (Sequence Type: action star, echoic onlooker, metaphor, and original event panel) design with the eight conditions counterbalanced into eight lists using a Latin Square Design.

As the inferential assessment scores showed that three strips from Experiment 1 were too hard to infer, these strips were also excluded from analyses in Experiment 2.

7.2. Participants

Participants were again recruited via social media and a participant pool available through Tilburg University, and all gave their consent to participate. The sample consisted of 70 participants with a mean age of 29.73 years ($SD = 10.98$, range: 17–61, 32 male, 35 female, 3 other). Although statistical power was not calculated a priori, a post hoc power analysis in G*Power again indicated sufficient power. A sample of 70 participants across eight conditions required F-values of above 2.03 to achieve a medium effect size of 0.25, which were met. The mean VLFI score was again average, at 12.24 ($SD = 9.48$, range: 1.6–34).

7.3. Procedure

The procedure was the same as for Experiment 1.

7.4. Data analysis

The same outlier removal process was used as in Experiment 1, resulting in removal of six participants, for a sample of 70 completed responses. Viewing times and comprehensibility ratings were averaged across items for each participant.

First, we conducted a 2 (Position: critical panel and critical panel +1) × 2 (Modality: unimodal and multimodal) × 4 (Sequence Type: action star, echoic onlooker, metaphor, and original event panel) factorial ANOVA for viewing times to examine the influence of position, the inclusion of a sound effect, and type of Peak. The data were further analyzed using a 2 (Modality: unimodal and multimodal) × 4 (Sequence Type: action star, echoic onlooker, metaphor, and original event panel) factorial ANOVA to look at viewing times at the Peak and subsequent panels separately, and the comprehensibility rating for the sequence. A regression again tested properties of the stimulus items, including the features as predictors. Additional correlations examined VLFI scores with the difference between viewing times

https://doi.org/10.1017/langcog.2022.19 Published online by Cambridge University Press

of inferential techniques and viewing times of the normal sequence. The data and analyses are accessible in an online data repository (https://doi.org/10.34894/DTBW7M).

8. Results

8.1. Viewing times

Viewing times were analyzed in a 2 (Position: critical panel and critical panel +1) × 2 (Modality: unimodal and multimodal) × 4 (Sequence Type: action star, echoic onlooker, metaphor, and original event panel) factorial ANOVA, which showed a main effect of position, $F(3, 1{,}104) = 160.49$, $p < 0.001$, $\eta\text{partial}_2 = 0.12$. As in Experiment 1, panels at the critical panel +1 position were viewed slower (M = 1,697.31, SD = 845.00) than those at the critical Peak panel position (M = 1,181.30, SD = 520.94). There was also a main effect of sequence type, $F(3, 1{,}104) = 11.83$, $p < 0.001$, $\eta\text{partial}_2 = 0.03$. Post hoc analyses indicated that panels in sequences with the original events (M = 1,314.20, SD = 541.19) were viewed faster than metaphors (M = 1,618.51, SD = 840.36) and echoic onlookers (M = 1,481.66, SD = 810.70). Moreover, sequences with action stars (M = 1,342.86, SD = 725.39) were viewed faster than those with metaphors. Finally, an interaction appeared between position and sequence type, $F(3, 1{,}104) = 12.81$, $p < 0.001$, $\eta\text{partial}_2 = 0.03$, which was explored further through separate analyses for each position.

At the critical Peak panel, viewing times were analyzed in a 2 (Modality: unimodal and multimodal) × 4 (Sequence Type: action star, echoic onlooker, metaphor, and original event panel) design, and differed across Sequence Type, $F(3, 552) = 12.80$, $p < 0.001$, $\eta\text{partial}_2 = 0.07$. As depicted in Fig. 6, action stars were viewed faster than echoic onlookers, metaphors, and original event panels (all $p < 0.001$). A main effect also appeared for Modality, $F(1, 552) = 8.18$, $p = 0.004$, $\eta\text{partial}_2 = 0.02$, because viewing times for multimodal panels were slower than unimodal panels. There was no interaction, $F(3, 552) = 0.22$, $p = 0.883$.

At the critical panel +1, a main effect of Sequence Type, $F(3, 552) = 12.14$, $p < 0.001$, $\eta\text{partial}_2 = 0.06$, arose because, as shown in Fig. 6, panels after original events were viewed faster than those after action stars, echoic onlookers, and metaphors (all $ps < 0.005$). There was no main effect for Modality, nor an interaction (all $p > 0.576$).

To explore the seemingly consistent difference across conditions evoked by the onomatopoeia, we calculated the mean difference between Modality conditions by subtracting the unimodal condition from the multimodal one. At the Peak panel, all viewing times to multimodal conditions increased by roughly 85–160 ms. An ANOVA tested the difference in viewing times at the Peak panel as the dependent variable, with Sequence Type as the independent variable. There was no main effect, $F(3, 280) = 0.48$, $p = 0.698$, suggesting no differences between sequence types. This consistent increase at the Peak panel hints at a uniform processing time required by the additional word across multimodal versions. For the subsequent panel, viewing times of multimodal action stars and original event panels increased by roughly 54 and 87 ms, respectively, whereas echoic onlookers and metaphors decreased by around 150 and 88 ms. Here too, no main effect was found ($p = 0.062$).

https://doi.org/10.1017/langcog.2022.19 Published online by Cambridge University Press

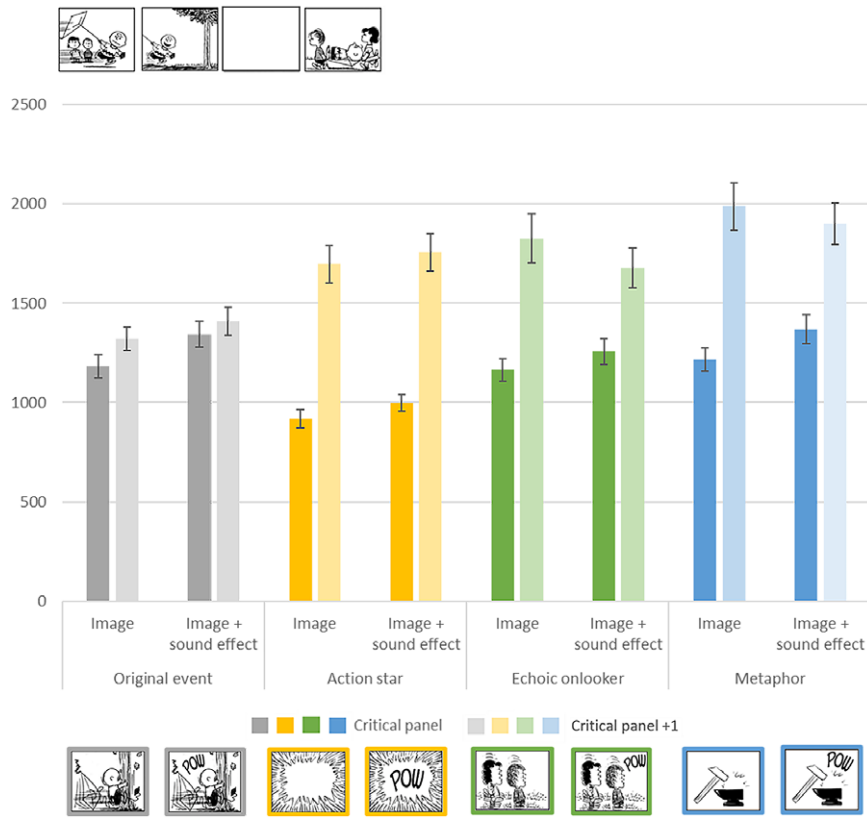

Fig. 6. Overview of viewing times at the critical Peak panel and subsequent panel for all eight sequence types; the error bars represent standard error.

8.2. Comprehensibility rating

A main effect of Sequence Type, $F(3, 552) = 49.62$, $p < 0.001$, $\eta\text{partial}_2 = 0.21$, reflected that sequences with the original Peaks were rated as most comprehensible (all $p < 0.001$), followed by action stars and then the echoic onlookers, which were both rated higher than metaphors (all $ps < 0.005$), as shown in Fig. 7. There was no main effect for Modality, nor an interaction (all $ps > 0.304$).

8.3. Features analysis

Here too, we first correlated the relevant features of [blend], [framing], and [explicit] against one another to test their relationship and found them to be valid predictors with a shared variance of 0.11 at most. General dominance and relative importance analyses can be found in the online repository.

At the critical Peak panel, the multiple regression for [blend], [framing], and [explicit] showed that the model explained 32.6% of the variance in the viewing times ($R^2 = 0.11$, $R^2_{\text{Adjusted}} = 0.09$, $F(3, 220) = 8.75$, $p < 0.001$). Table 3 reports the t-values and p-values of each feature, and Fig. 8 graphs the positivity or negativity of the

https://doi.org/10.1017/langcog.2022.19 Published online by Cambridge University Press

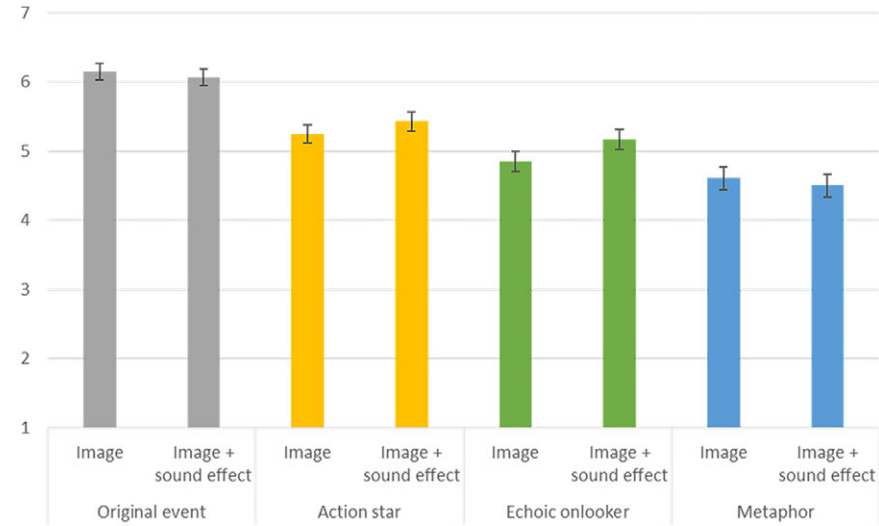

Fig. 7. Overview of the comprehensibility ratings for all eight sequence types; the error bars represent standard error.

Table 3. Overview of *t*-values and *p*-values of each feature per dependent variable

	Critical panel		Critical panel +1		Ratings	
Feature	*t*-value	*p*-value	*t*-value	*p*-value	*t*-value	*p*-value
Blend	4.14	<0.001	3.82	<0.001	−8.08	<0.001
Framing	0.74	0.458	3.43	<0.001	−5.06	<0.001
Explicit	3.88	<0.001	−2.12	0.035	2.37	0.019

standardized betas. Both [blend] and [explicit] predicted slower viewing times. At the critical panel +1, the features explained 33.2% of the variance in viewing times ($R^2 = 0.11$, $R^2_{\text{Adjusted}} = 0.10$, $F(3, 220) = 9.07$, $p < 0.001$). Here, [blend] and [framing] led to longer viewing times, whereas [explicit] predicted faster viewing times. For the comprehensibility ratings, the features explained 53.6% of the variance ($R^2 = 0.29$, $R^2_{\text{Adjusted}} = 0.28$, $F(3, 220) = 29.61$, $p < 0.001$). [Blend] and [framing] predicted lower ratings, and [explicit] predicted higher ratings.

As in Experiment 1, we also tested cloze probability and inference assessment as predictors, but again no relations emerged.

8.4. Visual language fluency

At the critical Peak panel, VLFI scores correlated with the differences between original events with multimodal action stars and with unimodal metaphors (respectively, $r(68) = 0.24$, $p = 0.040$, and $r(68) = 0.24$, $p = 0.041$). Smaller differences were found for experienced comic readers between original event panels and multimodal action stars, suggesting that these were read faster by experienced viewers, while expertise again led to longer viewing times for unimodal metaphors than original

https://doi.org/10.1017/langcog.2022.19 Published online by Cambridge University Press

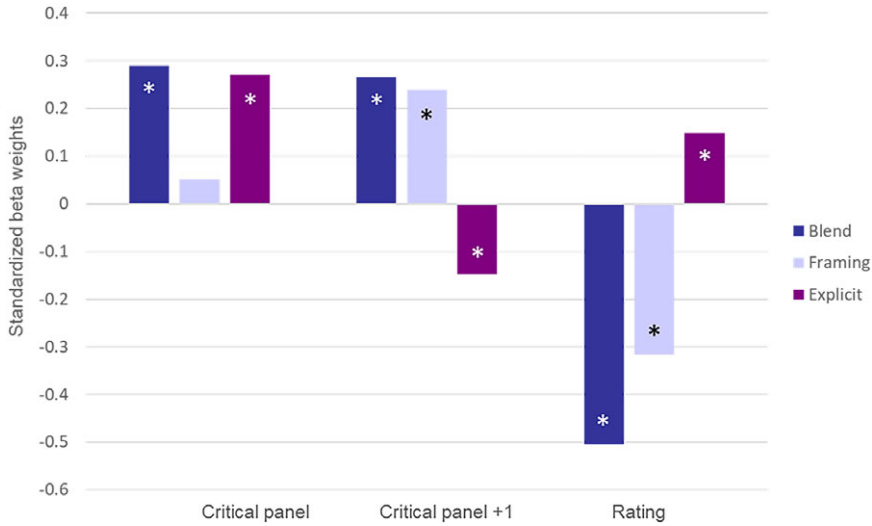

Fig. 8. Beta-weights from a regression showing the influence of different features on the viewing times and self-rated comprehension of the sequence.

event panels, as the mean difference score here was a positive value. At the critical panel +1, VLFI scores correlated with the difference between original event panels and action stars in both versions (multimodal: $r(68) = 0.30$, $p = 0.012$; unimodal: $r(68) = 0.31$, $p = 0.008$), and metaphors in both versions (multimodal: $r(68) = 0.25$, $p = 0.035$; unimodal: $r(68) = 0.28$, $p = 0.020$). In all cases, higher comic expertise (VLFI) correlated with slower viewing times for these sequence types' final panel relative to the original event panels.

9. Discussion

Experiment 2 combined inferential techniques to investigate the effect on inferential processing and comprehension, and the influence of the (combined) features. At the Peak, panels with minimal visual content (action stars and their multimodal versions) were viewed faster than the other techniques, similar to Experiment 1 (see also Cohn & Wittenberg, 2015). Moreover, all multimodal panels were processed longer than the unimodal versions, but this did not create differences at the subsequent panel. As in Experiment 1, all panels following an inferential technique required more time than panels following the original Peaks. These results indicate that adding explicit cues to inferential techniques does not facilitate inference resolution, but may increase processing demands at the Peak.

At the Peak, SPECT (Loschky et al., 2020) again well explains the results, since readers slowed down for panels combining techniques, and thus with more visual cues, relative to the unimodal Peaks. Multimodal sequence types had a noticeably consistent increase in viewing times, suggesting a similar number of extra fixations needed to extract the additional onomatopoeia, while little inference resolution occurred at this moment in the sequence. As in Experiment 1, the action stars were viewed the fastest, reinforcing the effect of visual complexity. While possibly

https://doi.org/10.1017/langcog.2022.19 Published online by Cambridge University Press

motivated by the visual cues alone, these results also aligned with the predictions of the [explicit] feature slowing down viewing times.

At the subsequent panel, longer viewing times appeared to all inferential techniques than the original event panel, again suggesting that the inference is resolved after the uninformative Peak. The added sound effect did not facilitate inference resolution, since no difference arose between the unimodal and multimodal inferential techniques. However, a trend arose that viewing times seemed to slightly increase for original panels and action stars when combined with sound effects and decreased for echoic onlookers and metaphors with sound effects. This supported [blend] and [framing] features aligning with slower viewing times seemingly more so than the increase in viewing times associated with additional explicitness. This trend in general may suggest that a more overt cue is not beneficial when combined with other techniques; possibly, such integration is more costly. It seems that the current experiment lacked power to reveal such a two-way interaction, which would be relevant for future research to test further.

Nevertheless, readers did process the sound effect, as shown at the Peak, but this explicit cue then had no apparent benefit for processing the subsequent panel. Potentially, sound effects are perceived as 'supplementary' elements to complement the action but with the visual events still carrying the most semantic weight (Cohn, 2016). As reinforced by the ratings, the onomatopoeia could be easily deleted with no consequences for the comprehensibility of the sequence. Thus, they appear not essential for deriving meaning here. On their own, as in Experiment 1, they facilitated inference resolution, so seemingly onomatopoeias have different semantic contributions based on how they are used.

In addition, no difference in comprehensibility ratings appeared for multimodal versus unimodal sequences. These ratings suggest that comprehensibility seems most informed by the inference resolution, which appeared not to be affected by the multimodality. As in Experiment 1 and previous work (Cohn & Kutas, 2015), the original event panels were rated as the most comprehensible. This aligns with explicitness predicting higher ratings. [Framing] and [blend] both predict low scores, concurring with low ratings for onlookers and metaphors. Action stars are rated relatively high, most likely due to being a familiar part of the visual lexicon of comics (Cohn, 2021; Cohn & Wittenberg, 2015).

In these contrasts, the features now predicted an even more recognizable trade-off, where more explicit cues at the Peak facilitate viewing times at the subsequent panel and ultimately higher comprehensibility ratings, as in previous studies (Cohn & Kutas, 2015; Cohn & Paczynski, 2013). This is also consistent with eye-movement studies showing that typicality affects how fast a bridging inference is constructed (Myers et al., 2000) and that predictive inferences facilitate processing when the following information aligns (Calvo, 2010; Calvo et al., 2001). Forward inferencing requires the context to be highly constraining, which has been tested with our inference assessment scores. Presumably, the explicit cue sets up a likely interpretation and if this is reflected in the final panel, processing of this event is facilitated. Moreover, as in Experiment 1, though not significant at the Peak, [framing] led to increased effort at the subsequent panel and lower ratings. Last, in these results, [blend] also affected the subsequent panel and ratings, but consistently negative rather than a reverse effect across panels.

https://doi.org/10.1017/langcog.2022.19 Published online by Cambridge University Press

10. General discussion

This paper presented two experiments comparing inferential climactic panels in visual narratives, expanding on previous work on inference which focused on omitting Peak information (Hutson et al., 2018; Magliano et al., 2016, 2017). Experiment 1 found that inferential techniques indeed differ in processing and comprehensibility, with visual complexity, explicitness, and framing emerging as contributing factors. Experiment 2 showed further that combining onomatopoeia may not necessarily clarify the missing event, despite being relatively easy to understand on its own. Across both studies, underlying features exerted competing influences on viewing times, but [explicit] and [framing] features consistently informed the processing of the subsequent panel and overall sequence comprehensibility.

Our primary finding was that inferential techniques motivate different processing. At both the Peak and subsequent panels, variations in viewing times arose across the techniques in both experiments. In combination with comprehensibility ratings, these results could support that some techniques were more successful in implying the unseen event than others. In Experiment 1, lower ratings aligned with longer viewing times at the panel following an inferential Peak. As inference resolution was hypothesized to occur at this panel, it implies that harder to interpret events led to readers understanding it less. In Experiment 2, panels following inferential techniques produced similar viewing times, but such sequences were rated distinctively. Thus, comprehensibility may not always align with the incremental panel-to-panel processing.

As with other studies, these results indicate that inference generation primarily occurs at the panel following inferential Peaks, despite such techniques operating specifically to omit events in different ways. Rather, at the Peaks, visual complexity exerted strong influence on viewing times, with less complex elements such as action stars and onomatopoeias processed faster. Such results are in line with SPECT's prediction that fewer cues to extract lead to fewer fixations (Loschky et al., 2020), and complement work examining eye-tracking of sequences with omitted information to generate bridging inferences (Hutson et al., 2018). There, an omitted event motivated a more intense search for cues at the final panel to facilitate generating the inference. As this study replicates the longer viewing times for panels after omitted events (Cohn & Wittenberg, 2015; Hutson et al., 2018; Magliano et al., 2016), it also implies that more fixations occur following when an inferential Peak is present, not just when events are omitted outright.

Although inferential Peaks are distinct, similarities persisted in their features at a more latent level. Analysis of these features suggest that they function like 'preference rules', exerting relative strengths. Yet, some noteworthy trade-off patterns appeared across our studies: explicitness consistently facilitated viewing times at the subsequent panel and enhanced comprehensibility, while [framing] slowed down viewing times and decreased ratings. In Experiment 2, the [explicit] effect was even more overt, stretching the pattern to the Peak as well. Here, longer processing at the Peak ultimately benefited processing and comprehensibility judgment afterward. This difference across experiments is likely motivated by more panels including explicit cues in the second set-up, since the additional cue was not only informative, but its mere presence naturally necessitated more eye fixations. This cascading effect of features of a Peak panel on the subsequent sequence suggested that such features at

https://doi.org/10.1017/langcog.2022.19 Published online by Cambridge University Press

least in some cases may extend beyond descriptive theoretical constructs, as posited in Cohn (2019). Despite these hints at the influence of features, the precise alchemy of featural interactions remains unknown, supported also by the varying influence of [blend] across experiments. Future work could continue to explore these features with other contrasts.

Overall, we investigated the consequences of using inferential Peak panels to elicit bridging inference rather than omitting the Peak entirely. While differences persisted across techniques, the explicitness and framing of the inferential Peak consistently informed their processing and comprehension. Inference always occurs within the context of a particular structure, and as shown here, this structure influences how inferences may be drawn. Thus, identifying and studying those patterns is essential for studying inference, beyond merely omitting events.

Data availability statement. The data that support the findings of this study are openly available in *Processing and understanding inferential techniques in visual narratives* at https://doi.org/10.34894/DTBW7M, V2.

References

Brisard, F., Frisson, S., & Sandra, D. (2001). Processing unfamiliar metaphors in a self-paced reading task. *Metaphor and Symbol*, 16(1–2), 87–108. https://doi.org/10.1080/10926488.2001.9678888

Calvo, M. G. (2010). The time course of predictive inferences depends on contextual constraints. *Language and Cognitive Processes*, 15, 293–319. https://doi.org/10.1080/016909600386066

Calvo, M. G., Meseguer, E., & Carreiras, M. (2001). Inferences about predictable events: Eye movements during reading. *Psychological Research*, 65(3), 158–169. https://doi.org/10.1007/s004260000050

Coderre, E. L., O'Donnell, E., O'Rourke, E., & Cohn, N. (2020). Predictability modulates neurocognitive semantic processing of non-verbal narratives. *Scientific Reports*, 10(1), 10326. https://doi.org/10.1038/s41598-020-66814-z

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

Cohn, N. (2013b). Visual narrative structure. *Cognitive Science*, 37(3), 413–452. https://doi.org/10.1111/cogs.12016

Cohn, N. (2014a). The architecture of visual narrative comprehension: The interaction of narrative structure and page layout in understanding comics. *Frontiers in Psychology*, 5, 680. https://doi.org/10.3389/fpsyg.2014.00680

Cohn, N. (2014b). You're a good structure, Charlie Brown: The distribution of narrative categories in comic strips. *Cognitive Science*, 38(7), 1317–1359. https://doi.org/10.1111/cogs.12116

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

Cohn, N. (2019). Being explicit about the implicit: Inference generating techniques in visual narrative. *Language and Cognition*, 11(1), 66–97. https://doi.org/10.1017/langcog.2019.6

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

Cohn, N. (2020b). Your brain on comics: A cognitive model of visual narrative comprehension. *Topics in Cognitive Science*, 12(1), 352–386. https://doi.org/10.1111/tops.12421

Cohn, N. (2021). A starring role for inference in the neurocognition of visual narratives. *Cognitive Research: Principles and Implications*, 6(1), 8. https://doi.org/10.1186/s41235-021-00270-9

Cohn, N. & Kutas, M. (2015). Getting a cue before getting a clue: Event-related potentials to inference in visual narrative comprehension. *Neuropsychologia*, 77, 267–278. https://doi.org/10.1016/j.neuropsychologia.2015.08.026

Cohn, N. & Paczynski, M. (2013). Prediction, events, and the advantage of agents: The processing of semantic roles in visual narrative. *Cognitive Psychology*, 67(3), 73–97. https://doi.org/10.1016/j.cogpsych.2013.07.002

Cohn, N. & Wittenberg, E. (2015). Action starring narratives and events: Structure and inference in visual narrative comprehension. *Journal of Cognitive Psychology*, 27(7), 812–828. https://doi.org/10.1080/20445911.2015.1051535

de Vries, C., Reijnierse, W. G., & Willems, R. M. (2018). Eye movements reveal readers' sensitivity to deliberate metaphors during narrative reading. *Scientific Study of Literature*, 8(1), 135–164. https://doi.org/10.1075/ssol.18008.vri

Dijk, T. A. V. & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press.

Fauconnier, G. & Turner, M. (2002). *The way we think: Conceptual blending and the mind's hidden complexities*. Basic Books.

Foulsham, T. & Cohn, N. (2020). Zooming in on visual narrative comprehension. *Memory & Cognition*, 49 (3), 451–466. https://doi.org/10.3758/s13421-020-01101-w

Frisson, S. & Pickering, M. J. (1999). The processing of metonymy: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(6), 1366–1383. https://doi.org/10.1037/0278-7393.25.6.1366

Frisson, S. & Pickering, M. J. (2007). The processing of familiar and novel senses of a word: Why reading Dickens is easy but reading Needham can be hard, 22(4), 595–613. https://doi.org/10.1080/01690960601017013

Gavaler, C. & Beavers, L. A. (2018). Clarifying closure. *Journal of Graphic Novels and Comics*, 11(2), 182–210. https://doi.org/10.1080/21504857.2018.1540441

Giora, R. (2003). *On our mind: Salience, context, and figurative language*. Oxford University Press.

Glucksberg, S. (2003). The psycholinguistics of metaphor. *Trends in Cognitive Sciences*, 7(2), 92–96. https://doi.org/10.1016/S1364-6613(02)00040-2

Glucksberg, S. & McGlone, M. (2001). *Understanding figurative language: From metaphor to idioms*. Oxford University Press.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101(3), 371–395. https://doi.org/10.1037/0033-295X.101.3.371

Guynes, S. A. (2014). Four-color sound: A Peircean semiotics of comic book onomatopoeia. *Public Journal of Semiotics*, 6(1), 58–72. https://doi.org/10.37693/pjos.2014.6.11916

Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2022). lab.js: A free, open, online study builder. *Behavior Research Methods*, 54, 556–573.

Huff, M., Meitz, T. G. K., & Papenmeier, F. (2014). Changes in situation models modulate processes of event perception in audiovisual narratives. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(5), 1377–1388. https://doi.org/10.1037/a0036780

Huff, M., Rosenfelder, D., Oberbeck, M., Merkt, M., Papenmeier, F., & Meitz, T. G. K. (2020). Cross-codal integration of bridging-event information in narrative understanding. *Memory & Cognition*, 48(6), 942–956. https://doi.org/10.3758/s13421-020-01039-z

Hutson, J. P., Magliano, J. P., & Loschky, L. C. (2018). Understanding moment-to-moment processing of visual narratives. *Cognitive Science*, 42(8), 2999–3033. https://doi.org/10.1111/cogs.12699

Jackendoff, R. & Lerdahl, F. (2006). The capacity for music: What is it, and what's special about it? *Cognition*, 100(1), 33–72. https://doi.org/10.1016/j.cognition.2005.11.005

Kosie, J. E. & Baldwin, D. (2019). Attentional profiles linked to event segmentation are robust to missing information. *Cognitive Research: Principles and Implications*, 4(1), 8. https://doi.org/10.1186/s41235-019-0157-4

Kuperberg, G. R., Paczynski, M., & Ditman, T. (2010). Establishing causal coherence across sentences: An ERP study. *Journal of Cognitive Neuroscience*, 23(5), 1230–1246. https://doi.org/10.1162/jocn.2010.21452

Lakoff, G. & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive Science*, 4, 195–208. https://doi.org/10.1016/S0364-0213(80)80017-6

Lerdahl, F., Jackendoff, R., & Slawson, W. (1985). A reply to Peel and Slawson's review of 'A generative theory of tonal music'. *Journal of Music Theory*, 29(1), 145. https://doi.org/10.2307/843373

Loschky, L. C., Larson, A. M., Smith, T. J., & Magliano, J. P. (2020). The scene perception & event comprehension theory (SPECT) applied to visual narratives. *Topics in Cognitive Science*, 12(1), 311–351. https://doi.org/10.1111/tops.12455

Magliano, J. P., Kopp, K., Higgs, K., & Rapp, D. N. (2017). Filling in the gaps: Memory implications for inferring missing content in graphic narratives. *Discourse Processes*, 54(8), 569–582. https://doi.org/10.1080/0163853X.2015.1136870

Magliano, J. P., Larson, A. M., Higgs, K., & Loschky, L. C. (2016). The relative roles of visuospatial and linguistic working memory systems in generating inferences during visual narrative comprehension. *Memory & Cognition*, 44(2), 207–219. https://doi.org/10.3758/s13421-015-0558-7

Magliano, J. P. & Zacks, J. M. (2011). The impact of continuity editing in narrative film on event segmentation. *Cognitive Science*, 35(8), 1489–1517. https://doi.org/10.1111/j.1551-6709.2011.01202.x

Manfredi, M., Cohn, N., De Araújo Andreoli, M., & Boggio, P. S. (2018). Listening beyond seeing: Event-related potentials to audiovisual processing in visual narrative. *Brain and Language*, 185, 1–8. https://doi.org/10.1016/j.bandl.2018.06.008

Manfredi, M., Cohn, N., & Kutas, M. (2017). When a hit sounds like a kiss: An electrophysiological exploration of semantic processing in visual narrative. *Brain and Language*, 169, 28–38. https://doi.org/10.1016/j.bandl.2017.02.001

McCloud, S. (1993). *Understanding comics: The invisible art*. Kitchen Sink Press.

Myers, J., Cook, A., Kambe, G., Mason, R., & O'Brien, E. (2000). Semantic and episodic effects on bridging inferences. *Discourse Processes*, 29, 179–199. https://doi.org/10.1207/S15326950dp2903_1

Ojha, A., Ervas, F., Gola, E., & Indurkhya, B. (2019). Similarities and differences between verbal and visual metaphor processing: An EEG study. *Multimodal Communication*, 8(2), 20190006. https://doi.org/10.1515/mc-2019-0006

Ortiz, M. J., Grima Murcia, M. D., & Fernandez, E. (2017). Brain processing of visual metaphors: An electrophysiological study. *Brain and Cognition*, 113, 117–124. https://doi.org/10.1016/j.bandc.2017.01.005

Papenmeier, F., Brockhoff, A., & Huff, M. (2019). Filling the gap despite full attention: The role of fast backward inferences for event completion. *Cognitive Research: Principles and Implications*, 4(1), 3. https://doi.org/10.1186/s41235-018-0151-2

Pratha, N. K., Avunjian, N., & Cohn, N. (2016). Pow, punch, pika, and chu: The structure of sound effects in genres of American comics and Japanese manga. *Multimodal Communication*, 5(2), 93–109. https://doi.org/10.1515/mc-2016-0017

Reid, V. M. & Striano, T. (2008). N400 involvement in the processing of action sequences. *Neuroscience Letters*, 433(2), 93–97. https://doi.org/10.1016/j.neulet.2007.12.066

Rundblad, G. & Annaz, D. (2010). Development of metaphor and metonymy comprehension: Receptive vocabulary and conceptual knowledge. *British Journal of Developmental Psychology*, 28(3), 547–563. https://doi.org/10.1348/026151009X454373

St. George, M., Mannes, S., & Hoffman, J. E. (1997). Individual differences in inference generation: An ERP analysis. *Journal of Cognitive Neuroscience*, 9(6), 776–787. https://doi.org/10.1162/jocn.1997.9.6.776

Strickland, B. & Keil, F. (2011). Event completion: Event based inferences distort memory in a matter of seconds. *Cognition*, 121(3), 409–415. https://doi.org/10.1016/j.cognition.2011.04.007

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, 22(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5

Yang, C. L., Perfetti, C. A., & Schmalhofer, F. (2007). Event-related potential indicators of text integration across sentence boundaries. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(1), 55–89. https://doi.org/10.1037/0278-7393.33.1.55

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind/brain perspective. *Psychological Bulletin*, 133(2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273

Zwaan, R. A. & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2), 162–185.

Cite this article: Klomberg, B. & Cohn, N. (2022). Picture perfect peaks: comprehension of inferential techniques in visual narratives *Language and Cognition* 14: 596–621. https://doi.org/10.1017/langcog.2022.19