Being explicit about the implicit: inference generating techniques in visual narrative*

NEIL COHN

Tilburg University

ABSTRACT

Inference has long been acknowledged as a key aspect of comprehending narratives of all kinds, be they verbal discourse or visual narratives like comics and films. While both theoretical and empirical evidence points towards such inference generation in sequential images, most of these approaches remain at a fairly broad level. Few approaches have detailed the specific cues and constructions used to signal such inferences in the first place. This paper thereby outlines several specific entrenched constructions that motivate a reader to generate inference. These techniques include connections motivated by the morphology of visual affixes like speech balloons and thought bubbles, the omission of certain narrative categories, and the substitution of narrative categories for certain classes of panels. These mechanisms all invoke specific combinatorial structures (morphology, narrative) that mismatch with the elicited semantics, and can be generalized by a set of shared descriptive features. By detailing specific constructions, this paper aims to push the study of inference in visual narratives to be explicit about when and why meaning is 'filled in' by a reader, while drawing connections to inference generation in other modalities.

KEYWORDS: visual language, narrative, inference, metaphor, comics, film.

---

1. Introduction

The ability to infer information that is not provided in a text has long characterized research on narrative. In work on visual narratives like comics, such inferencing has been highlighted as a primary feature, where readers 'fill in the gaps' and use their 'imagination' to understand a narrative sequence (Gavaler & Beavers, 2018; McCloud, 1993; Postema, 2013; Yus, 2008). However, research has often discussed inference by positing vague principles of connecting information across images (e.g., Groensteen, 2007; McCloud, 1993), or discussing the cognitive effects of making such inferences (Loschky, Hutson, Smith, Smith, & Magliano, 2018; Magliano, Larson, Higgs, & Loschky, 2015), yet rarely do studies outline explicit patterns that warrant such inferences in the first place. This paper will begin to fill *this* gap with regard to visual narratives, by outlining specific constructions that prompt a reader to generate inferences. It thereby aims to push the study of such inferences beyond generalities, as such specificity is crucial for understanding how the mind and brain comprehend visual narratives.

In the following, we will first discuss previous research on inference in language at the discourse and sentence level, and then connect this work to inference research in visual narratives. Following this, I will outline explicit inference generating techniques in both individual images and sequential images, and then conclude by presenting an analytical framework describing the features of these constructions.

1.1. DISCOURSE AND INFERENCE

In the study of discourse, inference generation and its cognitive correlates have been a central line of research. Though a reader engages with the surface presentation of a discourse, readers use this information to build a ***situation model*** of a scene – the comprehension of the 'situation' of a narrative's characters, location, events, etc. (Zwaan, 2004; Zwaan & Radvansky, 1998). This mental model is the understanding of the growing scene as it unfolds, including those not expressed overtly. Because the surface form might leave out various pieces of information, inferences by a reader 'fill in' information that may be absent from, unstated in, or connected across a discourse (Clark, 1977; Graesser, Singer, & Trabasso, 1994; McNamara & Magliano, 2009).

Inferences characterize nearly all semantic information possible in a discourse. Referential inference can connect units, like knowing that 'it' in a sentence refers to a previously mentioned object (anaphoric inference). Inference related to referential understanding can also arise from relations between objects used in actions or scenes, like knowing that being in snow involves wearing a jacket, even if left unstated (Metusalem et al., 2012).

67

Spatial inferences may also be made about implicit aspects of scenes' locations, such as tracking changes in spatial locations across a discourse (Zwaan & van Oostendorp, 1993). Metaphors or conceptual blends may also demand inference, such as how saying that *a surgeon is a butcher* implies a medical professional with a lack of skill, despite butchers being very skilled at their craft (Fauconnier & Turner, 1998; Lakoff & Johnson, 1980). Readers may also infer the internal states of characters, such as their goals, intentions, and emotions, both through characters' own or others' actions (Magliano, Taylor, & Kim, 2005; Trabasso, 2005). For example, why a character does something may not be apparent until after they do it, requiring this information to be inferred. Such inferences also relate to events more broadly, such as only knowing the result of an action, which sponsors an inference about what occurred prior to that result.

These types of backward-looking inferences are broadly classified as **bridging inferences** (Clark, 1977; Graesser et al., 1994; McNamara & Magliano, 2009), which 'fill in' unstated information based on prior and current information. In contrast, **predictive** or **forward inferences** foreshadow subsequent information (Magliano, Dijkstra, & Zwaan, 1996; McKoon & Ratcliff, 1986). As a situation model is built across a discourse, greater discontinuity will demand more updating of this mental model (McNamara & Magliano, 2009; Zwaan & Radvansky, 1998). Thus, because bridging inferences elicit an understanding that is not expressed, they require a growing situation model to be updated in order to comprehend the inferred information.

Inferences are not limited to the discourse level. Linguists have long debated about anaphora at the syntactic level (e.g., Chomsky, 1981) and ellipsis of information from sentence structure (Culicover & Jackendoff, 2005). In addition, some sentence structures involve 'enriched composition' whereby information is inferred or 'coerced', such as that *He began the book* implies reading or writing, or *She pounced until dawn* implies a repeated action, not a single sustained jump (Jackendoff, 1997; Pustejovsky, 1991). Thus, 'filling in' unstated information is a feature across levels of linguistic structure, and arises from mismatches in the correspondence between surface form and the construed meaning (Jackendoff, 1997, 2002). Here, we extend this notion to visual narratives to investigate how inference arises via mismatches in the correspondence between the *visual* form and meaning.

1.2. VISUAL NARRATIVES AND INFERENCE

Theories of inference in visual narratives have typically discussed bridging inferences, though not always using this terminology (Branigan, 1992;

Gavaler & Beavers, 2018; McCloud, 1993; Postema, 2013; Saraceni, 2001; Yus, 2008). This work has emphasized that readers 'fill in' information between panels, an image unit of a visual discourse. This is most prominently codified in McCloud's (1993) model of panel transitions, and the inferences generated by these juxtapositions of images (his term: 'closure'). Such transitions have been directly likened to coherence relationships between sentences (Saraceni, 2000, 2001; Stainbrook, 2003), where more inference is evoked by greater panel-to-panel discontinuity on dimensions like repeating information across panels or sharing a broader semantic field (Saraceni, 2000, 2001).

McCloud's transitions were posited as occurring only one at a time and non-incrementally (no partial transitions). Yet, semantic changes across juxtaposed panels are not 'all or nothing' transitions, and many dimensions might change across images concurrently (Cohn, Pederson, & Taylor, 2017; Gavaler & Beavers, 2018). Such meaningful shifts likely evoke mental model updating throughout a (visual) discourse, which may or may not reach the level of a full inference (Loschky et al., 2018). Indeed, mental model updating has been observed as an ongoing process at each image of a visual narrative, regardless of involving (in)congruity or significant bridging inferences (Cohn & Kutas, 2015). Incremental updating is a fairly general process, but bridging inferences are typically thought to incur more costs than these basic mappings, since they require a reader to alter a mental model given the absence of explicit information (Loschky et al., 2018; McNamara & Magliano, 2009).

At this point, experimental evidence supports that readers do generate inferences while comprehending a sequence of images. However, comparatively little discussion has been made of the *specific* mechanisms that warrant such inferences: what aspects of the *visual surface* allow for inferences to be generated in the *situation model*? We address this question first by discussing inference generating techniques within *individual* images, motivated by the structure of 'visual morphology', and then proceed to inference generation techniques across *sequences* of images. Finally, we will integrate these techniques into an analytical framework accounting for various features of inference generation.

2. Inference through morphology

Individual images can evoke inferences in several ways. Basic referential inferencing occurs through framing that may not show a full character or object. For example, when a panel border shows someone from only knees up, a fluent reader will know the character still has legs below the knee. This type of *part–whole metonymic relation* (synecdoche) evokes a type of referential inference.

69
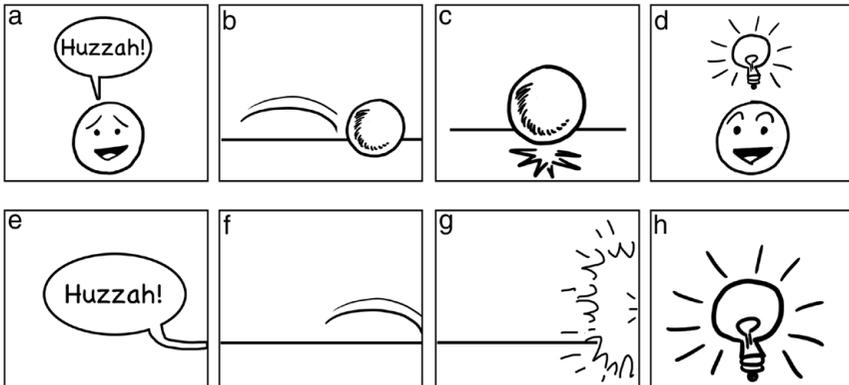
Fig. 1. Various visual morphology (a–c) that sponsors inferences when depicted off-panel (e–g) and other morphology that does not (d/h).

Framing can also interact with combinatorial qualities of 'visual morphology' (Cohn, 2013b) involved in non-representational aspects of images. Many visual 'morphemes', like speech balloons or motion lines, attach as affixes to 'stems', like a speaker or a mover (Cohn, 2013a, 2013b, 2018; Cohn, Murthy, & Foulsham, 2016; Forceville, 2011). These elements would be considered bound morphemes, as the affixes cannot stand alone and must attach to their stem. While the stem, like a speaker or a mover, can exist without the affix, omission of the stem might seem weird: bound morphemes like speech balloons or motion lines cannot float unattached to something (Cohn, 2018). This is analogous to language where affixes in English like *un-* or *-tion* cannot exist in a sentence without attaching to a stem word.

Nevertheless, visual stems can be omitted under constrained circumstances, particularly through panel framing, and because of their bound relationship, such cases instigate inference of the missing stem. Consider Figure 1a–c, which depicts several visual bound morphemes: (a) a 'carrier' of a speech balloon, (b) a motion line, and (c) an impact star. All of these elements attach to a stem of an object (a face or ball). Now consider Figure 1e–g, where these same morphemes extend to the panel borders, and the stem has been placed 'off-panel'. Though the stem is no longer depicted, it remains inferred through the indexicality of the morpheme: they connect to *something* off-panel. There is a stem, even if we do not see it.

Thus, because of the combinatorial link between bound morphemes and their stems, stems can be inferred even if off-panel or occluded. This can allow for creative story-telling, like the famous *Alpha Flight #6* by John Byrne (Marvel, 1984), titled *Snowblind*, which used five pages of only white panels with carriers and text under the conceit that the story took place in a snowstorm. These pages depicted no stems, but their meaning

70

could be inferred through the morphological relationship of these affixes. Comparable conditions occur when characters are in the dark, with all-black panels showing only carriers like speech balloons or thought bubbles.

Similar inferencing occurs when carriers attach to inanimate objects that may *contain* the stem. For example, a thought bubble above Snoopy's doghouse would not be interpreted as the doghouse 'thinking', but rather of Snoopy thinking while he is *inside* of it. Other examples include balloons or bubbles attaching to buildings, cars, or other such enclosures. This inference may arise because balloons and bubbles license an animate stem (Cohn, 2013a), which combines with the knowledge that the objects are containers (buildings, vehicles). So, if a carrier attaches to an inanimate container, an inference extends to the stem inside it, rather than the object itself construed as speaking or thinking.

Nevertheless, not all bound morphemes generate this type of inference. 'Upfixes' are visual affixes that float above characters' heads, such as light bulbs to mean inspiration (Figure 1d), or hearts to mean love (Cohn, 2013b; Cohn et al., 2016). Upfixes are highly constrained in their positions (Cohn, 2013b; Cohn et al., 2016; Forceville, 2011), and yet do not use an indexical relation to their stems like those in Figure 1a–c. Thus, leaving the stem for an upfix outside the panel does not create a demand for inference like other visual morphemes, meaning that Figure 1h just seems like an actual light bulb, rather than inspiration for a head that cannot be seen.

In sum, inference in single panels arises from framing of basic referential information or combinatorial aspects of visual morphology. As we will see, both framing and bound morphemes interact with other inference generating techniques across panel *sequences*.

3. Inference in sequential images

As mentioned, discussion of inferences in visual narratives like comics often revolves around recognizing content missing from a sequence. However, omitted meanings are not uniform, and may depend on what type of information is deleted. **Visual Narrative Grammar** (VNG) is a model of the narrative structure of sequential images, describing the system that organizes and presents semantic information to a comprehender (Cohn, 2013c, 2015). This theory posits that several independent components contribute to a visual narrative sequence in a **parallel architecture** similar to and integrated with the architecture of language (Cohn, 2015, 2016b; Jackendoff, 2002). This allows one structure or another to change, while others remain intact. These different structures include those related to the meaning of a visual sequence (spatial, referential, and event structures) and those related to its presentation (graphic and narrative structures).

71
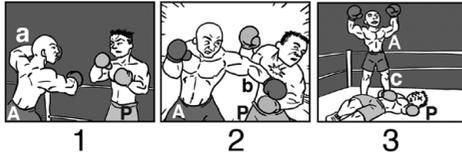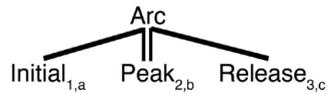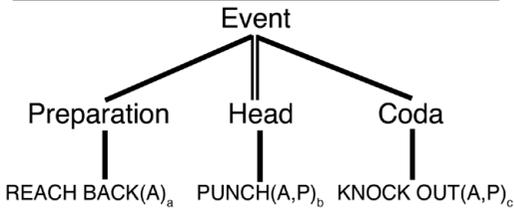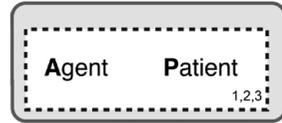
Fig. 2. Parallel structures involved in sequential image understanding, including those presenting the narrative (graphic, narrative) and those about meanings (event, and spatial/referential structure). Correspondences between structures are indicated for panels (numbers), event elements (lower-case subscripts), and referential entities (upper-case letters).

Figure 2 depicts this parallel architecture for a short visual sequence (Cohn, 2015). Here, the *graphic structure* depicts the surface form of the visual utterance (lines, shapes). This representation links to the meanings of the unfurling actions and events of the sequence (*event structures*), the characters involved (*referential structure*), and the spatial organization of a scene (*spatial structure*), these latter two here collapsed together for simplicity. Finally, this meaning is organized sequentially by the *narrative structure*. These components are interfaced in various ways, here notated with subscripts for event information (lower-case letters), referential entities (upper-case letters), and panels (numbers).

VNG posits that narrative structure is built analogously to syntax at the sentence level, but corresponding to a higher level of information structure (semantics) closer to a discourse. These grammatical properties include basic categorical roles for units, which are organized into hierarchic constituent structures. VNG follows contemporary models of *construction grammar* which posit that grammatical schemas are stored in memory, along with their licensed mappings to semantic structures (Culicover & Jackendoff, 2005; Goldberg, 1995). Thus, narrative schemas encoded in memory map to semantics (event, spatial, referential structures) which feed into a situation model of the ongoing constructed visual discourse. Ultimately, the surface information of the narrative grammar is posited to fade from memory, while the meaning constructed in the situation model persists.

72

TABLE 1. *Two basic constructional patterns in Visual Narrative Grammar. X denotes a variable that can be filled by any category from the canonical schema*

a) Canonical schema:	[$_{\text{Phase X}}$ (Establisher)-(Initial)-(Prolongation)-Peak-(Release)]
b) Conjunction schema:	[$_{\text{Phase X}}$ X$_1$ - X$_2$ -... X$_n$]

The canonical narrative schema (Table 1a) orders categories into a basic sequence. An *Establisher* may start a sequence, setting up the characters involved in an interaction. This may be followed by an *Initial*, where the events and interactions begin, prototypically with a preparatory action. A *Prolongation* may then extend this anticipatory state before the climax in the *Peak,* which is often a completion or interruption of an action. A subsequent *Release* may then dissipate the tension of the interaction, often with the coda of an event. As noted with parentheses in Table 1a, all of the categories of the canonical narrative schema are optional, except the Peak, which is the motivating 'head' of the constituent (indicated by double-bar lines). However, as discussed below, omission of narrative categories is more complex than this, including for Peaks.

These categories manifest in a fairly basic sequence in Figure 2, where each narrative category maps to prototypical event structures (lower-case letter subscripts): a boxer reaches back his arm (preparation; Initial), punches another boxer (head event; Peak), and then stands triumphantly having knocked out his opponent (Coda; Release). This preparation–head–coda structure characterizes the basic elements of discrete events (Jackendoff, 2007), as suggested by manipulating or omitting aspects of this structure (Cohn, Paczynski, & Kutas, 2017; Lasher, 1981; Strickland & Keil, 2011), and such structure can embed recursively to form hierarchic event structures (Jackendoff, 2007; Zacks & Tversky, 2001). In its most basic form, the canonical narrative arc maps isomorphically to event structures, but more complex narratives belie this mapping.

In addition, the mapping between narrative and spatial structures specifies how much information occurs in each panel. Here, both characters are shown in all panels. Thus, the dotted lines drawn around both Agent and Patient represent that panels divide up this spatial structure by enclosing both entities.

Narrative categories can apply to both panels and to constituents composed of several panels, making the canonical narrative schema inherently recursive. Because of this, narrative structures can extend to higher levels, such as to whole plotlines. However, the mechanisms of inference generation described throughout can be characterized within single constituents, so our discussion will remain at this simpler level.
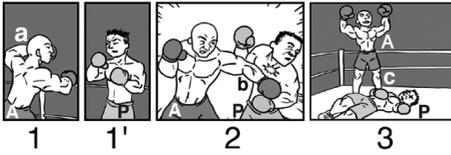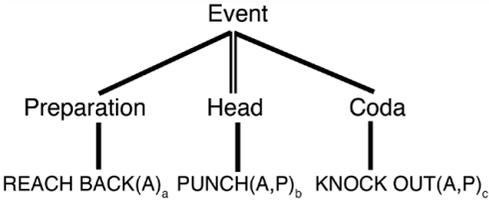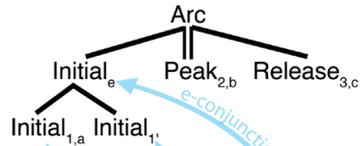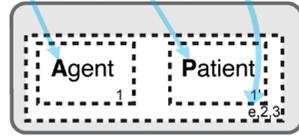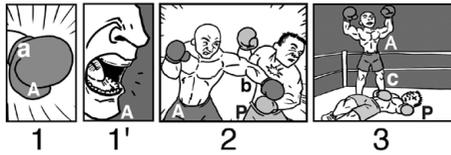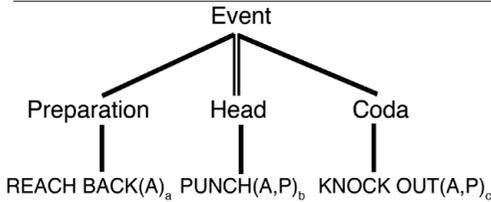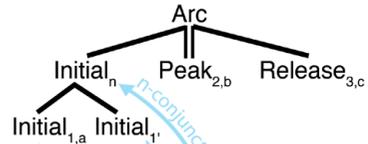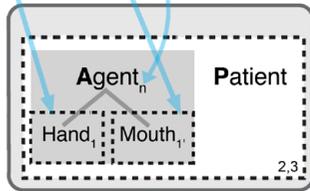
Fig. 3. Mappings between narrative and semantics for (a) Environmental-Conjunction and (b) Entity-Conjunction.

Additional patterns in VNG elaborate on the canonical schema. Consider the conjunction schema (Table 1b), which repeats a single narrative category ('X') within a constituent of that same category (Cohn, 2013c, 2015). Thus, multiple Initials in a conjunction would repeat within a single Initial constituent. Conjunction can involve various mappings to semantics, as in Figure 3.

Figure 3a breaks apart the first two panels of Figure 2. Now, each character occupies their own panel, but the location shared by these characters needs to be inferred. This inference is the equivalent of the first panel of Figure 2. This pattern is described as Environmental-Conjunction (E-Conjunction), because both panels function as Initials (like the single panel in Figure 2),

now conjoined in an Initial constituent (Figure 3a's narrative structure). These panels then map to a spatial inference enclosing both characters in a common environment. Thus, in spatial structure, the Agent and Patient each have a dotted line indicating the panels where they appear individually (panel 1 and 1′), while another dotted line encompasses them both for the panels where they appear together (panels 2 and 3) and for the inferred space created from conjoining panels 1 and 1′ (notated as 'e' for 'environment').

Neurocognitive research has implicated both combinatorial processing (narrative grammar) and updating processes (situation model) involved in Environmental-Conjunction (Cohn & Kutas, 2017). In addition, corpus research has suggested that Japanese manga use Environmental-Conjunction more often than American and European comics (Cohn, in press), and concomitantly, frequency of readership of manga modulates the brain response to Environmental-Conjunction (Cohn & Kutas, 2017). Specifically, frequent manga readers used more combinatorial processing, while less frequent manga readers used more updating processes. Thus, familiarity with specific narrative constructions can vary their processing, beyond just general mechanisms.

Another part–whole inference occurs in Figure 3b. Both panels again act as Initials, but here show parts of a single character, rather than parts of a scene. This creates an Entity-Conjunction (N-Conjunction) (Figure 3b) where a character or object ('entity') must be inferred from its parts (this inference notated as 'n'). Both the narrative and event structures remain the same between Figure 3a and 3b; but they differ in the spatial/referential structure linking to the conjunction. Thus, conjunction can involve part–whole inferences of environments and entities.

3.1. ELLIPSES OF NARRATIVE CATEGORIES

Given these structural properties, we can now identify how omissions function in a narrative sequence. Some panels can delete easier than others (Cohn, 2014; Magliano, Kopp, Higgs, & Rapp, 2017). Establishers can acceptably be deleted from drawn sequences with little consequence (Cohn, 2014), and this appears to persist in other modalities like film (Kraft, Cantor, & Gottdiener, 1991) and motion graphics (Barnes, 2017). Initials and Peaks are generally unacceptable to delete, though when they are, their semantic content may be inferred, as discussed below (Cohn & Kutas, 2015; Cohn & Wittenberg, 2015; Magliano et al., 2015). Finally, a sequence may make sense without a Release, but it will create a fairly abrupt ending. These tendencies have been supported by studies where participants choose panels to delete from a sequence, or recognize which panels may have been deleted (Cohn, 2014). Additional studies have found viewing times increase to panels following dispreferred omissions,
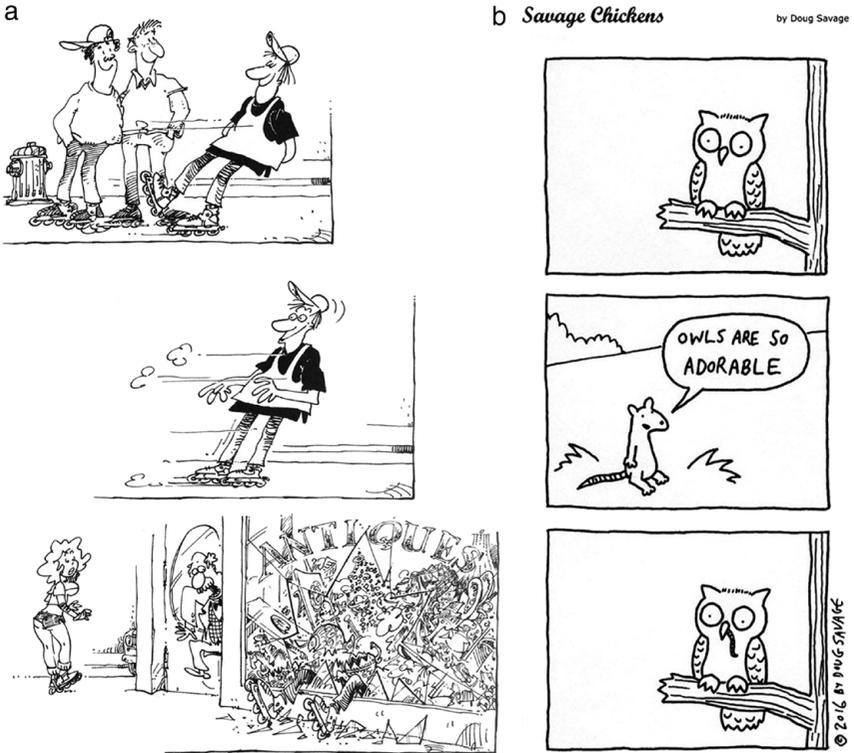
Fig. 4. Two strips which both omit Peak panels which thereby cause the inference of climactic events prior to the final panel. (a) *Actions Speak* © 2002 Sergio Aragonés. (b) Savage Chickens © 2016 Doug Savage (www.savagechickens.com).

as categorized in a complementary analytical framework (Magliano et al., 2017). Overall, omission of different narrative categories can vary the comprehensibility of the sequence. The variance of these ellipses illustrates that inference motivated by omission is not uniform (McCloud, 1993), but is contingent on (and can be characterized by) the narrative structure itself.

3.2. PEAK DROP

As discussed, deleting Peaks from a sequence lowers its comprehensibility and increases the inference required to understand it. Nevertheless, Peaks can be omitted intentionally with constrained regularity. Consider the two strips in Figure 4. Figure 4a shows a man skating backwards, who is then distracted by something in the second panel (cued by wide eyes and open lips). The final panel shows him lying in a broken window of an antique store, while the distracting element is revealed as an attractive woman. This final

76

image shows the aftermath of an event (its coda), but the sequence never shows the climatic Peak: where he crashes into the window. Similarly, Figure 4b depicts an owl in the first panel, with a mouse commenting how cute it is in the second (both Initials, united via Environmental-Conjunction). The final panel shows the mouse's tail in the owl's mouth. Again, the primary event (the owl eating the mouse) is left inferred, while the coda is depicted.

These examples use **Peak Drop**, the purposeful omission of a Peak which generates an inference for an omitted causal event from the combination of its preparatory (Initial) and aftermath states (Release). Because a Peak is essential to a narrative sequence, and such cases appear to be systematic, this may be a special construction which omits the Peak from the canonical narrative schema. This would differ from simple elision of Peaks from a canonical schema, in that Peak Drop invokes a schematic pattern that structurally intends for Peak to be missing and the events to be inferred.

A Peak Drop Construction thus uses a narrative phase with no Peak, which maps to semantic information of a causative event (head) where the resulting state (coda) maps to the Release. Figure 5a illustrates this relationship within VNG. The combination of the preparatory and resulting events thus generate the inference about the missing event (Strickland & Keil, 2011), though it is absent in the narrative structure. Such omission of a Peak elicits an updating process, evident in slower self-paced viewing times to panels following omitted Peaks, despite fairly high ratings of comprehensibility (Cohn & Wittenberg, 2015; Hutson, Magliano, & Loschky, 2018; Magliano et al., 2015). Updating is also suggested by visual search processes aiming to find inferential cues to clarify the preceding absent event information (Hutson et al., 2018). Such inference generation appears to overlap with linguistic working memory systems, implying that such mechanisms may be domain-general (Magliano et al., 2015).

Nevertheless, the claim is that a *narrative* Peak is indeed missing from a sequence, not just that its semantic content is inferred. Such an interpretation is supported by studies of visual narrative where omitting Peak panels show a cost relative to those that contain a Peak, as indicated through longer viewing times (Cohn & Wittenberg, 2015) or larger neural responses indicative of reanalysis (Cohn & Kutas, 2015), even when both involve inference generation. Readers cognitively recognize a missing *narrative* Peak, above and beyond the inference of the missing semantic information (Cohn & Kutas, 2015).

3.3. SUPPLETIVE PANELS

Besides deletion, panels may be *replaced* with inference-generating information. **Suppletive panels** are a class of panels – often systematic – which replace information at various levels of explicitness. Suppletive panels mostly express three types of concepts: general events, fights, and transformations.
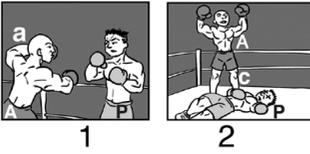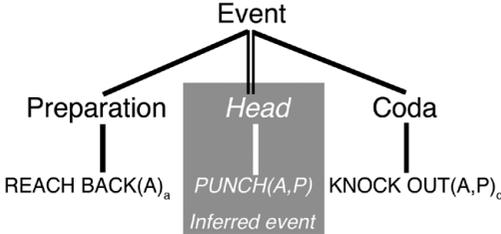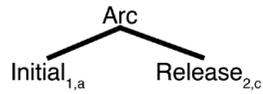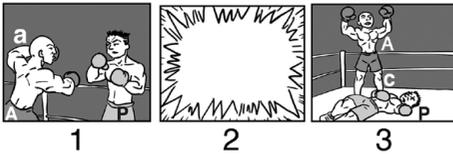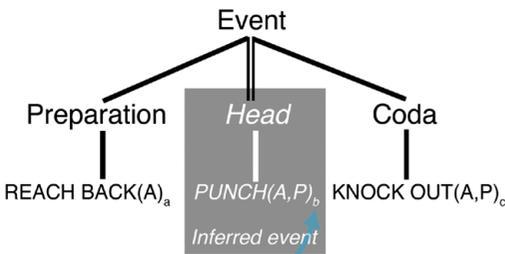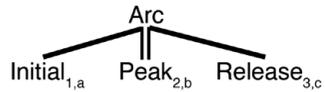
Fig. 5. Missing event information generated by (a) Peak Drop where the Peak panel is omitted, and (b) replacement of a Peak with an action star panel.

The most general suppletive panel is an action star, as in Figure 5b, which depicts an impact star (Figure 1c), blown up to the size of a full panel (Cohn, 2013b). This star shape is homonymous with a carrier showing loudness, i.e., a jagged speech balloon or sound effect (Cohn, 2013b; Forceville, Veale, & Feyaerts, 2010), and indeed, action stars often co-occur with text, like *Pow!* (Manfredi, Cohn, & Kutas, 2017). Thus, action stars are morphologically ambiguous as either a blown-up impact star or a jagged-edged carrier.

Nevertheless, action stars express a sudden and potentially impactful event, without depicting what that event actually is. In Figure 5b, the event structure of the action star is encoded as a sudden and violent unspecified event (notated as a variable: 'X') between the Agent (A) and Patient (P), and thus an inference is required to infer its meaning (i.e., punch, connected to the action star through the subscript 'b'). Text like *Pow!* or *Smack!* could provide greater specificity to this ambiguity (discussed below). With or without text, a reader must infer the missing events, pulling from information in the prior and subsequent panels, as in Peak Drop. Indeed, as with omitting a Peak event, readers view panels following an action star slower than following an explicit depiction of an event, implying that the longer time is caused by inferential processing (Cohn & Wittenberg, 2015).

Despite their impoverished semantics, action stars play a well-formed narrative role in the sequence. They replace Peak panels (Figure 5b), but less suitably replace other categories. Action stars are a sort of 'pro-Peak' in the way that *pro*nouns act as nouns; they satisfy the categorical role, but lack semantic specificity. This structural role is supported by findings that readers spend less time viewing an action star compared to a fully blank panel at the Peak position, despite the blank being less visually rich (Cohn & Wittenberg, 2015). Thus, action stars facilitate viewing times more than simply omitting event information, thereby suggesting a sequential role.

Other suppletive panels are more explicit. Fighting is replaced by fight clouds (Figure 6a), which typically show a large cloud, a tornado, or a jumble of lines, along with body parts, clothes, and/or other morphemes (spirals, stars, lightning bolts) emerging from the fray. Though action stars and fight clouds may both replace combat, they imply different dimensions of 'aspectuality' – i.e., how much time they convey (Cohn, 2013b). Action stars are a punctive event, implying a single momentary action. You cannot infer multiple punches or an exchange of punches in an action star. In contrast, fight clouds imply durative events, occurring across a span of time. Here, there must be multiple actions (hits, kicks, bites, etc.) and cannot be only a single action. Thus, both panels replace actions, but evoke subtly different inferences about those events. Nevertheless, like action stars, fight clouds as full panels can function as Peaks, signaling a high arousal event.

A *love cloud* also uses a cloud shape, but to imply sex instead of violence. This may be fairly saccharine, with hearts emerging (instead of the stars, like a fight cloud), as in Figure 6b, or with more salacious body parts sticking out. The overall intent of fight and love clouds are similar: rather than show the durative events directly, they become implied through a suppletive panel.

Various depictions also substitute for transformations, subsumed within a class of ***transformative suppletion panels*** (TSPs). For example, Figure 6c

Fig. 6. Suppletions of panels for specific events with (a) a fight cloud for fighting and (b) a love cloud for sex, by Dwayne Godwin and Jorge Cham, and a poof cloud transformation in *Rickety Skitch and the Gelatinous Goo* © Ben Costa and James Parks. Also, (d) a climactic panel (Peak) replaced by a panel solely with text from *Quantum and Woody Must Die* © Steve Lieber.

replaces a poof cloud for a depiction of the imp transforming to look like a gnome. Here, we infer that the imp transformed, despite seeing only the prior and end states; the poof cloud provides a climactic unit for the transformation, relieving the artist from showing the process overtly. TSPs have several allomorphic depictions, with slight variance in manner implied between clouds, blobs, stars, tornadoes, and other graphic features. Like action stars, TSPs imply a sudden punctive event, despite that transformations may extend across a duration of time. They also play narrative roles as Peaks.

80

A multimodal suppletion occurs when text replaces a panel, particularly onomatopoeia (*Pow!*) or a 'descriptive' (*Punch!*) sound effect (Guynes, 2014; Manfredi et al., 2017; Pratha, Avunjian, & Cohn, 2016). Consider Figure 6d, where a woman defends herself with a stapler against a monster. The climactic stapling moment is not shown, replaced instead with the sound effect *Pa-chunk!* This onomatopoeia replaces, and thereby demands inference of, the (Peak) event. Text substitution often accompanies other suppletions, like action stars and TSPs (Figure 6c).

Text-substitution elicits inference both from the sound effect and its surrounding context. The text provides event cues as a 'sound emission' (Goldberg & Jackendoff, 2004), whereby the sound resulting from an event indexes the actual event. In language, similar inferences arise from constructions like *Tom belched his way out of the room* to mean that Tom *went out* of the room while belching (Goldberg & Jackendoff, 2004). Further, if the text mismatches the visual context of the preceding or subsequent images (example in Figure 6d, text of *Smooch!*), it will appear incongruous (Manfredi et al., 2017). Thus, as with other suppletive devices, both content and context motivate the inference of the undepicted events.

Text-substitution thus connects a sound effect to the unseen event. This text leaves both the carrier (i.e., 'balloon') and the stem (the thing generating the sound) undepicted (Cohn, 2013a). This suppletion goes beyond the off-panel inference created by a tail pointing to the border of a panel (Figure 1e), instead omitting everything *but* the contents of the carrier. Again, the morphological connection between the text and its stem facilitates the inference, even if the actual carrier is not depicted (Cohn, 2013a).

Text-substitution also typically functions as a Peak, with its cues coming exclusively from the verbal, not visual modality (Manfredi et al., 2017). Cross-modal substitution of a grammatical category (narrative, syntax) occurs between all modalities – i.e., text into image sequences, images into text sequences, gesture into speech, etc. (Cohn, 2016b).

3.4. MENTAL-SPACE EVOKING PANELS

Panels can also imply events and actions through metonymy or metaphor, where depicted information if understood literally would be incongruous. However, in context, such panels evoke larger semantic frames or connect mental spaces to create conceptual blends or metaphors to form a coherent interpretation (Fauconnier & Turner, 1998; Forceville, 2016).

First, panels might show something related to an event, without showing the event itself. These replacements thus depict a ***metonymic*** relationship to the unseen events (Cohn, 2010; Joue et al., 2018; Kowalewski, 2018; Kukkonen, 2008). Consider Figure 7a, where a man in bed turns off the light.

81

Fig. 7. Panels which use mappings between mental spaces for (a) metonymy of a light being turned off in *One Night* by Tym Godek, and (b) metaphors for a sex scene in panels 1, 3, and 5; Deadpool is © Marvel Comics.

We are not shown an image of his hand turning off an actual lamp, but rather we see a darkened light bulb in full. The darkened light bulb is metonymic for the lamp's light being turned off, relative to the brightness of the prior panel and the darkness of the subsequent one. Other metonymies might depict doves flying or a statue crying instead of depicting a murderous event.

Panels might also use ***metaphors*** to imply unseen events. Metaphors abound in visual narratives, often embedded into their visual morphology, such as steam coming from characters' ears to depict anger as if the head was a pressurized container (Cohn, 2018; Forceville, 2016; Lakoff & Johnson, 1980). Yet, metaphors can also insinuate events (Cohn, 2010; Gavaler & Beavers, 2018). Consider Figure 7b, which comes from a full page where numerous panels use metaphors for sex (train through a tunnel, volcano erupting, etc.). Another metaphor would be a scene where someone becomes mad, then a panel of a tea kettle boiling over, then the other person with a black eye. The kettle would thus invoke the metaphor of anger as fluid in a pressurized container (Forceville, 2016; Lakoff & Johnson, 1980).

Figure 8 illustrates metaphor panels in the parallel architecture of Visual Narrative Grammar. The Peak has been replaced with a panel depicting an explosion. This panel should be incongruous, featuring none of the characters or situation involved in the sequence. However, if understood metaphorically, the events of an explosion become mapped to the event of a punch (subscript 'm'),

82
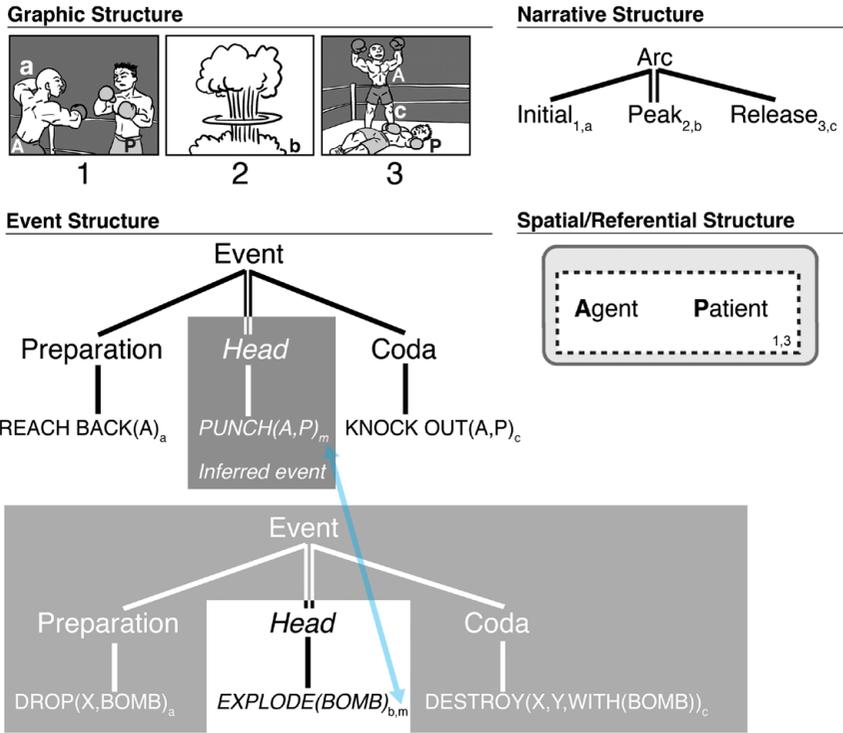
Fig. 8. Metaphoric mapping of seemingly incongruous event to the primary event in a sequence of images.

evoking parallels between the puncher's weapon (fist/bomb) destroying its target (boxer/bombed target). These correspondences arise from the inferences drawn between the conceptual spaces of the punching and bombing events.

As suppletive panels, metonymic and metaphoric panels exist outside the storyworld of the visual narrative. However, these mental space evoking relations can also incorporate elements from the narrative scene, which will be discussed below.

3.5. ONLOOKERS

Beyond omissions or replacements, inference can also occur by diverting focus to something else in a scene, like showing a person observing an event instead of the event itself. Such **onlookers** may vary in how explicitly they suggest what actions or events they watch. A *Passive Onlooker* is the least explicit voyeur, showing characters observing an action off-panel, but with a neutral unexpressive facial expression. Typically, the panel after the Passive

83
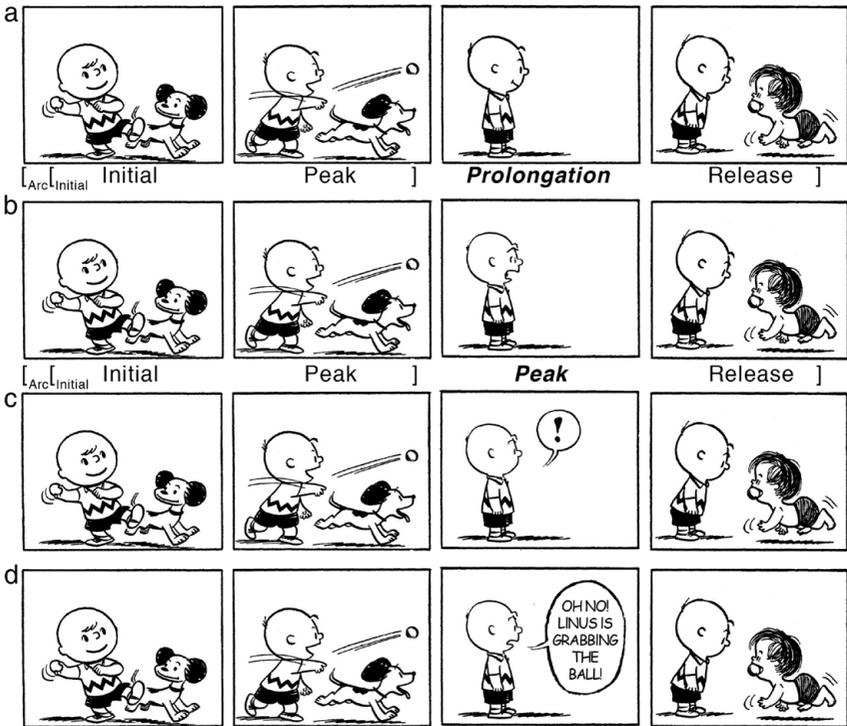
Fig. 9. Onlooker panels in a slightly manipulated *Peanuts* strip, where (a) depicts Charlie as a Passive Onlooker, (b) depicts him as a Reactive Onlooker, (c) depicts him with an exclamatory cue of a carrier, and (d) adds descriptive narration. *Peanuts* is © Peanuts Worldwide LLC.

Onlooker reveals an aftermath of what occurs out of the reader's view. In Figure 9a, Charlie Brown is the Passive Onlooker in panel 3, only to reveal at panel 4 that he watched Linus retrieve the ball, not Snoopy. Because the Passive Onlooker gives no cues about the events taking place, the final panel demands a bridging inference to fill in this missing information (Cohn & Kutas, 2015).

This final image is a narrative Release, as it shows the punchline or aftermath of an event. But what of the Passive Onlooker? Here, ambiguity remains as to whether the unseen event happens *at* this panel, watched by the onlooker, or if it happens *between* this and the final panel. If the climactic event occurs simultaneous to the onlooker, it would imply a Peak, but would only be recognized as such at the subsequent Release. Indeed, the semantic cues in this panel do not signal anything about a climactic event and, in fact, gives few clues about events at all. Thus, given its semantic cues, panel 3 of Figure 9a is more characteristic of a Prolongation, extending the prior state, and the subsequent Release implies that this sequence uses Peak Drop.

84

Variations to onlookers hint more about an unseen event. A *Reactive Onlooker* has an emotional facial expression that signals some action occurs off-panel. For example, Figure 9b alters 9a only by making Charlie's facial expression surprised. Though Reactive Onlookers' emotions generally suggest a viewed event (looking surprised, exasperated, grossed out, happy, etc.), this expression gives no clues about the actual event. This might align with the 'Spielberg face' in film, a film shot where an onlooker is shown with eyes wide in wonderment at an off-camera event (Fortunato, 2014).

Unlike Passive Onlookers, Reactive Onlookers signal a climactic event, and thus are narrative Peaks. A Reactive Onlooker is modeled in Figure 10a, with a Peak of a surprised referee, cued by the facial expression and exclamation mark that he is reacting to the off-panel event *at this moment*. His event (below, subscript b) is that of 'seeing', but what he sees is the whole event of the Agent and Patient (α). This index binds to the inferred event (the punch), meaning that he sees this whole event (α), while the reader does not.

Both Passive and Reactive Onlookers require inference to understand the unseen event, and thus draw from context in the previous and subsequent panels. A fairly explicit Initial may allow a strong inference at the Reactive Onlooker, such as when a person shoots a basketball in an Initial, followed by a Peak of the onlooker crowd cheering. The actual event (ball through the hoop) would not be depicted, but the inference is highly constrained. An event with less clear forecasting would thus generate inference at a subsequent panel to the onlooker. As stated, whereas the Passive Onlooker provides no cues of an off-panel event at that moment, a Reactive Onlooker signals that the inferential event occurs at that narrative moment, as in Figure 9a and 9b. Thus, Reactive Onlookers trigger the necessity for an inference prior to Passive Onlookers in the flow of the narrative sequence (Cohn & Kutas, 2015).

This contrast in cues between Passive Onlookers and Reactive Onlookers has consequences on their narrative structure. Because Passive Onlookers lack cues, they remain categorically ambiguous (likely a Prolongation), while cues in Reactive Onlookers can signal a Peak, except the climactic event is known to remain off-panel. Neurocognitive research has supported this interpretation because Reactive Onlookers (like Figure 9b) evoke greater neural responses consistent with mental model updating than Passive Onlookers (like Figure 9a), suggesting that these cues provided more explicit event information (Cohn & Kutas, 2015). Yet, at the subsequent panel, greater updating processes were suggested for Passive Onlookers than Reactive Onlookers. Indeed, neural responses to panels following Reactive Onlookers did not differ from panels following Peaks which explicitly showed the climactic events. Thus, despite requiring the *same inference* as Passive Onlookers, neural responses to panels after Reactive Onlookers patterned with those of regular Peaks, thereby suggesting they had a similar structural role.
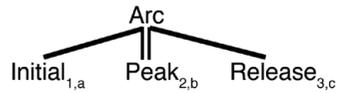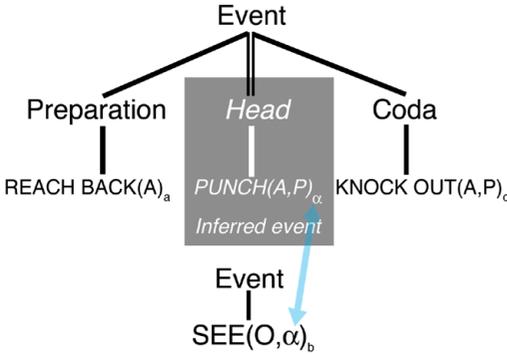
85
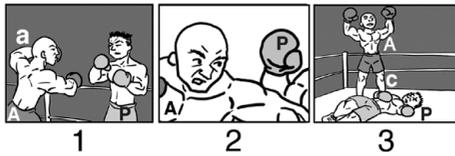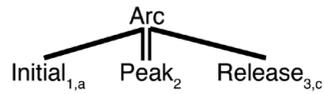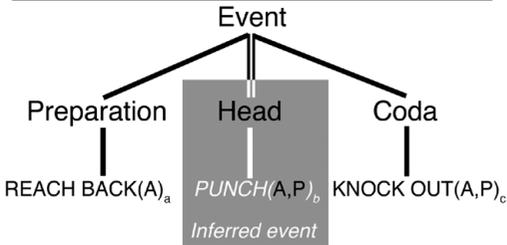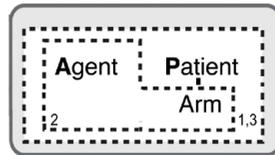
Fig. 10. Missing event information generated by (a) a Reactive Onlooker where the Peak panel depicts a voyeur of the actions, and (b) selective framing of non-essential event information.

Onlookers can become more expressive if their body posture fully implies the off-panel events. *Echoic Onlookers'* postures imitate actions taken by characters out of view. Consider Figure 11, where a character, Rob, attempts to pick up on women at a bar. His friend, Ramon, watches the encounter.

Fig. 11. Sequence with an Echoic Onlooker in the third panel. *Butternutsquash* © Ramon Perez and Rob Coughler.

In panel 2 (an Initial), Ramon looks apprehensive, as a Reactive Onlooker, with Rob implied by the tail of the balloon pointing to the panel's edge. In panel 3, the Peak, Ramon grabs his crotch and winces – suggesting that his friend has been kicked below the belt (reinforced by the text-substitutions: *Smack, pow, punt!*). Thus, Ramon in panel 3 is an Echoic Onlooker, 'echoing' the actions undergone by Rob off-panel. Echoic Onlookers are the most explicit type of onlooker, as their actions fully cue the inferred event.

Additional *exclamatory cues* can make onlookers more expressive. As in Figure 9c, adding upfixes to onlookers can provide non-bodily cues that signal an off-panel event. These include exclamation marks (!) and question marks (?), either floating as upfixes above a head, or placed in carriers (balloons or bubbles). Exclamatory cues of meaningful balloons or bubbles might say 'Gasp!' or some other general surprised statement by an onlooker. Like the bodily cues in Reactive or Echoic Onlookers, exclamatory cues signal the onlooker's perception of an off-panel event at that moment. Such cues can motivate the panel as a Peak, and, as discussed above, can elicit different brain responses than panels with Passive Onlookers (Cohn & Kutas, 2015).

Finally, an onlooker might convey more than a vague sense of surprise like with exclamatory cues, instead being completely explicit about the unseen events. *Descriptive narration* by an onlooker provides a play-by-play of what they see happening. In Figure 9d Charlie, the onlooker, now explicitly describes the unseen event, thereby making the final panel fairly unrevealing. Here, the event is not depicted *visually*, but it is overtly stated in the text. Thus, a comprehender would need to infer the visual properties of an event, but would need little inference from the holistic multimodal interaction.

3.6. SELECTIVE FRAMING

Other inference generating techniques may show elements in a scene related to an event, while not showing the actual event. Such cases use *selective framing*

of non-central aspects of an event, while still depicting the entities involved. Consider Figure 12a, where a man threatens the woman, only in panel 3 for her to grab his groin. Instead of showing her action, this panel selectively frames his pained expression in response to it. Similar relations are modeled in Figure 10b, where most of the structure remains the same as Figure 2, only now the panel windows a narrower amount of information (as reflected in spatial structure), which now makes the event in the Peak less salient.

Research with eye-tracking has indeed suggested that certain areas of a panel are more informative for the depicted events than others. Viewers tend to fixate event-informative regions (Hutson et al., 2018), whether the narrative sequence is coherent or incoherent (Foulsham, Wybrow, & Cohn, 2016), though fixations may be more dispersed in incoherent sequences. When framed as a whole panel, this fixated information would be similar to the first panel in Figure 12b, which shows only the primary action (hand pulling out a plug) without any of the surrounding details. Selective framing would thus window 'non-fixated' information where their indicative quality pushes a reader to infer the events.

Examples of selective framing include a sex scene depicting curled toes or an orgasmic face instead of the explicit sex act, or in a fight scene, a pale exasperated face instead of showing a person being stabbed, shot, etc. (as in Figure 12a). Throughout, the key visual information cuing the event would be missing. In some ways, onlookers share this windowing quality of selective framing by showing a person watching the events, rather than essential actors themselves. While selective framing can replace an explicit Peak, constrained windowing of information can apply to all panels in the narrative (example Figure 12b).

Other selective framing can window something in the scene other than the events, while retaining a semantic connection to them. *Metonymic selective framing* shows elements of a scene related to an event instead of the entities or cues involved in the event. Figure 12b shows a bullet casing flying in the air with smoke to indicate the firing of a gun. The gun is not shown firing, but the bullet discharge metonymically indicates the firing as a selectively framed part of the gunshot action. Other metonymic selective framing might include blood splattering without showing the wound itself, showing bubbles underwater instead of a person drowning, or, for the recurring boxing example, showing a mouthpiece flying through the air as the Peak.

The Peaks in Figure 12a–c also motivate inference through onomatopoeia. *Text selective framing* uses speech balloons, thought bubbles, or sound effects to indicate an unseen event. This is similar to the text-substitution onomatopoeia, but here may not depict a 'sound effect' (such as someone screaming 'Nooooo!' while killed off-panel) and may not fully replace a Peak.

88

Fig. 12. Peak panels (a) implying a man being grabbed in the groin; *Battle Chasers* © Joe Madureira, (b) a gunshot through the metonymic selective framing with a bullet casing and the accompanying emergent text (*Blam!*); *The Creech* © Greg Capullo, and (c) implying a skull being cracked through the crashing of a pumpkin; *Watchmen* © DC Comics.

89

Finally, *metaphoric selective framing* can use elements in a scene to imply other actions. Consider Figure 12c from *Watchmen*. The crashing pumpkin metaphorically stands in for the crushed head of the character in the first panel (Gavaler, 2017), accompanied by the off-panel affixing text of 'HHUUGGHH'. The pumpkin appears in the background of the first panel of Figure 12c, largely to set up its metaphorical function in the subsequent Peak. Metonymic and metaphoric selective framing differ from metonymic or metaphoric suppletions (discussed above) by whether the elements exist in the actual scene. Metonymies and metaphors involving selective framing show elements that exist within the scene itself and relate to the primary events, while suppletions shift to depict abstract information outside the storyworld.

4. Underlying features of inference generating techniques

Given these constructions for generating inference, it's worth highlighting underlying similarities between these techniques. These 'features' can descriptively tease apart the characteristics of inference generating situations, but whether they characterize psychological contributions to comprehenders' understanding is an open, empirical question. For these purposes they are useful for describing and analyzing the characteristics of these techniques. At least seven variables can characterize the inferential constructions discussed here:

- Substitutive: Panel depicts something in lieu of depicting the actual event.
- Framing: Panel frames certain elements from a scene instead of others.
- Explicit: Elements contain information about an event (e.g., action star is not explicit, since it is about generic events, but fight cloud is explicit, because it is just about fights).
- Blend: Depiction requires mapping between mental spaces to be understood (metaphoric, metonymic).
- Arousal: Depictions are high arousal (events) rather than low arousal (states or nothing).
- Affixation: Elements use a morphological affix.
- Text: Elements are in the verbal modality rather than the visual modality.

Table 2 shows the underlying 'features' for the various inference generating techniques, ordered from those requiring more inference (non-informative) to those requiring less inference (informative), based approximately on how much constructions omit or overtly provide information. A Peak Drop requires the most inference, because it is a full omission, while many substitutive

90

TABLE 2. *Inference generating techniques in visual language along with their various features. They are listed along a scale from non-informative (requiring more inference) to informative (requiring less inference).*

Inference generating technique	Text	Framing	Substitutive	Explicit	Blend	Arousal	Affix
Peak Drop							
Passive Onlooker		X	X				
Exclamatory Cues Onlooker	X	X	X			X	X
Reactive Onlooker		X	X			X	
Action star			X			X	X
Transformational suppletive panel			X			X	X
Fight/love cloud			X	X		X	X
Metonymic panels			X		X		
Metaphoric panel			X		X	X	
Echoic Onlooker		X	X	X		X	
Text-Substitution	X		X	X		X	X
Metonymic selective framing		X		X	X	X	
Metaphoric selective framing		X		X	X	X	
Text selective framing	X	X	X	X		X	X
Selective framing		X		X		X	
Environmental/ Entity-Conjunction		X		X			
Descriptive Text Onlooker	X		X	X		X	X

techniques provide some information, but not explicitly (e.g., Passive Onlookers are less informative than Echoic Onlookers) and selective framing depicts aspects of a scene itself. Nevertheless, the relative inferential strengths of these constructions remain an empirical question. Prior research has investigated some of these contrasts, such as comparisons between deletion of different narrative states (Cohn, 2014; Magliano et al., 2017), between action stars and Peak Drop (Cohn & Wittenberg, 2015), and between Passive and Reactive Onlookers through simple inclusion or omission of cues (Cohn & Kutas, 2015). However, careful manipulation can test the relative inferential demand(s) of different techniques, and explore whether such underlying features may manifest as psychological constructs.

Table 2 orders these techniques by explicitness *in isolation*, but such cues can combine, as most all naturalistic examples throughout attest. As described, text-substitution of onomatopoeia often accompanies action stars or selective framing techniques, while onlookers often use exclamatory cues. Some combinations between cues might be less acceptable though. An onomatopoeia for a punctive event (*Pow!*) alongside a durative fight cloud

91

would conflict in terms of temporal aspect. Similarly, an explicit onomatopoeia describing an event (*Pow!*) would conflict with a metaphoric panel (tea kettle boiling). In general, inferences involving blended mental spaces are less likely to combine with other elements, since both mental spaces would be simultaneously activated, and thereby conflict. In contrast, metonymic panels should effectively combine with text, as in Figure 12b, and similarly with the Echoic Onlooker (also technically metonymic) in Figure 11.

Combinations of too many explicit cues may create a strange sequence. For example, descriptive text with an Echoic Onlooker might be so informative as to reduce the effectiveness of an inference. Inference generating techniques designed to be less informative, and thus too many sources of information would flout that minimalism. Optimal combinations seem to balance cues suggesting an inference with their explicitness, thereby echoing Grice's (1967) maxim of quantity to give only as much information as needed, but no more.

Informativeness can also tie to a reader's experience. Theorists of both visual and verbal narrative have posited that narratives which force the reader to 'interact' with their content are more engaging (Herman, 2009; McCloud, 1993; Zwaan, 2004), and inference generating techniques sponsor such interactions. However, speculatively, the informativeness of a visual narrative may relate to reader enjoyment like an inverted U-curve (e.g., Berlyne, 1971). Visual narratives explicit about everything will be highly informative but may be less engaging, while an under-informative sequence demanding too much inference may also be less enjoyable because of being hard to comprehend. Thus, a balance between explicitness and inference generation may be most pleasurable for a reader. While most work has examined how inference generation interacts with comprehension, this dimension of reader engagement would be fruitful for further research.

Reader engagement may be one of several reasons that creators of visual narratives use these inference generating techniques. Some techniques may facilitate economy of representation, such as transformational suppletive panels providing a simple alternative for drawing the incremental parts of a character's transformation. Others may arise for discretion or censorship, such as not showing violence or sex (like action stars, onlookers, or love clouds), while others still might facilitate comedy (like onlookers in Figures 4, 9, and 11), playfulness or creativity (as in metonymy or metaphor), or evocativeness (such as a Reactive Onlooker viewing a particularly gruesome act). Finally, they may manipulate the narrative structure, increasing tension with more units (conjunction), or compressing a scene by omitting narrative categories. Authors no doubt balance these constraints while creating visual narratives, and exploring these functional constraints can also be an avenue for future research.

5. Conclusion

This paper has described various techniques for generating inference in visual narratives, framed within Visual Narrative Grammar. These patterns are argued to be 'lexicalized' constructions stored in the long-term memory of creators and readers of comics. This survey is surely not exhaustive, and many other inference generating techniques likely will be identified in future work. Further, corpus work can indicate whether these (and other) patterns may vary cross-culturally in visual narrative systems, and how such patterns may have emerged or changed over time. Such efforts are particularly important given that exposure to visual narrative constructions appears to modulate their processing (Cohn & Kutas, 2017).

In addition, these constructions may arise in modalities outside drawn visual narratives, and indeed similar notions appear in the literature on film. Narrative grammar has been applied to film (Amini, Riche, Lee, Hurter, & Irani, 2015; Cohn, 2016a), motion graphics (Barnes, 2017), and discourse (Kosara, 2017), and insofar as the principles of this model are cross-modal, so too should be such inference generating techniques. For example, many stimuli eliciting bridging inferences in experiments within the discourse literature use what is identified here as Peak Drop, albeit not framed with reference to a narrative structure (e.g., Kuperberg, Paczynski, & Ditman, 2011; Singer, Halldorson, Lear, & Andrusiak, 1992). Characterizing whether these particular inference generating techniques apply across modalities, and/or whether we could identify other patterns, would be fruitful for future research.

It is important to stress that this work aimed to identify *specific* patterns in visual narratives based on their structural properties. Visual sequences are stitched together from constructional patterns (Cohn, 2013c, 2015), parallel to syntactic constructions which build sentences (Culicover & Jackendoff, 2005). Within this, specific patterns and techniques generate inferences out of the interface between narrative and meaning. Thus, to understand the structure and comprehension of visual narratives, we must detail these specific patterns, rather than making broad proclamations about general principles that operate across sequences, inference included.

GRAPHIC REFERENCES

Aragonés, S. (2002). *Actions Speak*. Dark Horse Comics.
Asmus, J., & S. Lieber, S. (2015). *Quantum and Woody Must Die* #1. Valiant Comics.
Capullo, G. (1997). *The Creech* #1. Image Comics.
Costa, B., & Parks, J. (2017). *Rickety Skitch and the Gelatinous Goo*. Knopf.
Duggan, G., Posehn, B., & Koblish, S. (2012). *Deadpool* #13. Marvel Comics.
Godek, T. (2006). *One Night*. Originally posted on 20 March 2006 <http://www.yellowlight.scratchspace.net/comics/onenight/onenight.html>.

Godwin, D., & Cham, J. (2014). Might as well face it, you're addicted to love. *Scientific American Mind* online. Originally posted 14 February 2014. Online <https://dwaynegodwin.tumblr.com/image/84481552842>.

Madureira, J. (1998). *Battle Chasers* #3. Image Comics.

Moore, A., & Gibbons, D. (1987). *Watchmen*. DC Comics.

Perez, R., & Coughler, R. (2003). *Butternutsquash*. Originally posted 28 May 2003 <www.butternutsquash.net>.

Savage, D. (2016). Chickenowls. *Savage Chickens*. Originally posted on 30 May 2016 <www.savagechickens.com>.

REFERENCES

Amini, F., Riche, N. H., Lee, B., Hurter, C., & Irani, P. (2015). Understanding data videos: looking at narrative visualization through the cinematography lens. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 1459–1468). New York: ACM.

Barnes, S. (2017). Studies in the efficacy of motion graphics: the impact of narrative structure on exposition. *Digital Journalism*, **5**(10), 1–21.

Berlyne, D. E. (1971). *Aesthetics and psychobiology*. New York: Appleton-Century-Crofts.

Branigan, E. (1992). *Narrative comprehension and film*. London: Routledge.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Clark, H. H. (1977). Bridging. In P. N. Johnson-Laird & P. C. Wason (Eds.), *Thinking: readings in cognitive science* (pp. 243–263). Cambridge: Cambridge University Press.

Cohn, N. (2010). Extra! Extra! Semantics in comics! The conceptual structure of *Chicago Tribune* advertisements. *Journal of Pragmatics*, **42**(11), 3138–3146.

Cohn, N. (2013a). Beyond speech balloons and thought bubbles: the integration of text and image. *Semiotica*, **2013**(197), 35–63.

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

Cohn, N. (2013c). Visual narrative structure. *Cognitive Science*, **37**(3), 413–452.

Cohn, N. (2014). You're a good structure, Charlie Brown: the distribution of narrative categories in comic strips. *Cognitive Science*, **38**(7), 1317–1359.

Cohn, N. (2015). Narrative conjunction's junction function: the interface of narrative grammar and semantics in sequential images. *Journal of Pragmatics*, **88**, 105–132.

Cohn, N. (2016a). From visual narrative grammar to filmic narrative grammar: the narrative structure of static and moving images. In J. Wildfeuer & J. A. Bateman (Eds.), *Film text analysis: new perspectives on the analysis of filmic meaning*. (pp. 94–117). London: Routledge.

Cohn, N. (2016b). A multimodal parallel architecture: a cognitive framework for multimodal interactions. *Cognition*, **146**, 304–323.

Cohn, N. (2018). Combinatorial morphology in visual languages. In G. Booij (Ed.), *The construction of words: advances in construction morphology* (pp. 175–199). London: Springer.

Cohn, N. (in press). Structural complexity in visual narratives: theory, brains, and cross-cultural diversity. In Grishakova, M. & Poulaki, M. (Eds.), *Narrative complexity and media: experiential and cognitive interfaces* (pp. 174–99). Lincoln, NE: University of Nebraska Press.

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.

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**(27), 1–13.

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

Cohn, N., Paczynski, M., & Kutas, M. (2017). Not so secret agents: event-related potentials to semantic roles in visual event comprehension. *Brain and Cognition*, **119**, 1–9.

94

Cohn, N., Pederson, K., & Taylor, R. (2017). A picture is worth more words over time: multimodality and narrative structure across eight decades of American superhero comics. *Multimodal Communication*, **6**(1), 19–37.

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.

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

Fauconnier, G., & Turner, M. (1998). Conceptual integration networks. *Cognitive Science*, **22**(2), 133–187.

Forceville, C. (2011). Pictorial runes in *Tintin and the Picaros*. *Journal of Pragmatics*, **43**(3), 875–890.

Forceville, C. (2016). Conceptual metaphor theory, blending theory, and other cognitivist perspectives on comics. In N. Cohn (Ed.), *The visual narrative reader* (pp. 89–114). London: Bloomsbury.

Forceville, C., Veale, T., & Feyaerts, K. (2010). Balloonics: the visuals of balloons in comics. In J. Goggin & D. Hassler-Forest (Eds.), *The rise and reason of comics and graphic literature: critical essays on the form* (pp. 56–73). Jefferson, NC: McFarland & Company.

Fortunato, J. (2014). The gaze and the Spielberg face: Steven Spielberg's application of Lacan's mirror stage and audience response. *Visual Communication Quarterly*, **21**(1), 40–53.

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

Gavaler, C. (2017). *Superhero comics*. London: Bloomsbury.

Gavaler, C., & Beavers, L. A. (2018). Clarifying closure. *Journal of Graphic Novels and Comics*, 1–30. Retrieved from <https://doi.org/10.1080/21504857.2018.1540441>.

Goldberg, A. E. (1995). *Constructions: a Construction Grammar approach to argument structure*. Chicago, IL: University of Chicago Press.

Goldberg, A. E., & Jackendoff, R. (2004). The English resultative as a family of constructions. *Language*, **80**(3), 532–568.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, **101**(3), 371–395.

Grice, H. P. (1967). William James lectures: logic and conversation. In P. Cole & J. L. Morgan (Eds.), *1975: Syntax and Semantics 3* (pp. 41–58). New York: Academic Press.

Groensteen, T. (2007). *The system of comics* (trans. B. Beaty & N. Nguyen). Jackson, MS: University of Mississippi Press.

Guynes, S. A. (2014). Four-color sound: a Peircean semiotics of comic book onomatopoeia. *Public Journal of Semiotics*, **6**(1), 58–72.

Herman, D. (2009). *Basic elements of narrative*. Chichester: Wiley-Blackwell.

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

Jackendoff, R. (1997). *The architecture of the language faculty*. Cambridge, MA: MIT Press.

Jackendoff, R. (2002). *Foundations of language: brain, meaning, grammar, evolution*. Oxford: Oxford University Press.

Jackendoff, R. (2007). *Language, consciousness, culture: essays on mental structure (Jean Nicod lectures)*. Cambridge, MA: MIT Press.

Joue, G., Boven, L., Willmes, K., Evola, V., Demenescu, L. R., Hassemer, J., … Habel, U. (2018). Handling or being the concept: an fMRI study on metonymy representations in coverbal gestures. *Neuropsychologia*, **109**, 232–244.

Kosara, R. (2017). *An argument structure for data stories*. Paper presented at the Proceedings of the Eurographics Conference on Visualization (EuroVis) 2017, Barcelona, Spain (pp. 31–35). doi:10.2312/eurovisshort.20171129

Kowalewski, H. (2018). Heart is for love: cognitive salience and visual metonymies in comics. *The Comics Grid: Journal of Comics Scholarship*, **8**, doi:http://doi.org/10.16995/cg.117

Kraft, R. N., Cantor, P., & Gottdiener, C. (1991). The coherence of visual narratives. *Communication Research*, **18**(5), 601–616.

95

Kukkonen, K. (2008). Beyond language: metaphor and metonymy in comics storytelling. *English Language Notes*, **46**(2), 89–98.

Kuperberg, G. R., Paczynski, M., & Ditman, T. (2011). Establishing causal coherence across sentences: an ERP study. *Journal of Cognitive Neuroscience*, **23**(5), 1230–1246.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Lasher, M. D. (1981). The cognitive representation of an event involving human motion. *Cognitive Psychology*, **13**(3), 391–406.

Loschky, L. C., Hutson, J. P., Smith, M. E., Smith, T. J., & Magliano, J. (2018). Viewing static visual narratives through the lens of the Scene Perception and Event Comprehension Theory (SPECT). In A. Dunst, J. Laubrock, & J. Wildfeuer (Eds.), *Empirical comics research: digital, multimodal, and cognitive methods* (pp. 217–238). London: Routledge.

Magliano, J. P., Dijkstra, K., & Zwaan, R. A. (1996). Generating predictive inferences while viewing a movie. *Discourse Processes*, **22**, 199–224.

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.

Magliano, J. P., Larson, A. M., Higgs, K., & Loschky, L. C. (2015). The relative roles of visuospatial and linguistic working memory systems in generating inferences during visual narrative comprehension. *Memory & Cognition*, **44**(2), 207–219.

Magliano, J. P., Taylor, H., & Kim, H.-J. J. (2005). When goals collide: monitoring the goals of multiple characters. *Memory and Cognition*, **33**(8), 1357–1367.

Manfredi, M., Cohn, N., & Kutas, M. (2017). When a hit sounds like a kiss: an electro-physiological exploration of semantic processing in visual narrative. *Brain and Language*, **169**, 28–38.

McCloud, S. (1993). *Understanding comics: the invisible art*. New York: Harper Collins.

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **12**(1), 82–91.

McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of Learning and Motivation*, **51**, 297–384.

Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., & Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. *Journal of Memory and Language*, **66**(4), 545–567.

Postema, B. (2013). *Narrative structure in comics: making sense of fragments*. Rochester: RIT Press.

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.

Pustejovsky, J. (1991). The syntax of event structure. *Cognition*, **41**, 47–81.

Saraceni, M. (2000). *Language beyond language: comics as verbo-visual texts*. Unpublished doctoral dissertation, University of Nottingham.

Saraceni, M. (2001). Relatedness: aspects of textual connectivity in comics. In J. Baetens (Ed.), *The graphic novel* (pp. 167–179). Leuven: Leuven University Press.

Singer, M., Halldorson, M., Lear, J. C., & Andrusiak, P. (1992). Validation of causal bridging inferences in discourse understanding. *Journal of Memory and Language*, **31**(4), 507–524.

Stainbrook, E. J. (2003). *Reading comics: a theoretical analysis of textuality and discourse in the comics medium*. Unpublished doctoral dissertation, Indiana University of Pennsylvania.

Strickland, B., & Keil, F. (2011). Event completion: event based inferences distort memory in a matter of seconds. *Cognition*, **121**(3), 409–415.

Trabasso, T. (2005). Goal plans of action and inferences during comprehension of narratives. *Discourse Processes*, **39**(2), 129–164.

Yus, F. (2008). Inferring from comics: a multi-stage account. *Quaderns de Filologia. Estudis de Comunicacio*, **3**, 223–249.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, **127**(1), 3–21.

Zwaan, R. A. (2004). The immersed experiencer: toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 35–62). New York: Academic Press.

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

Zwaan, R. A., & van Oostendorp, H. (1993). Do readers construct spatial representations in naturalistic story comprehension? *Discourse Processes*, **16**, 125–143.