Acta Psychologica 164 (2016) 157–164

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

The pieces fit: Constituent structure and global coherence of visual narrative in RSVP

Carl Erick Hagmann [a,*], Neil Cohn [b]

[a] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, United States
[b] Department of Cognitive Science, University of California, San Diego, La Jolla, CA, United States

ARTICLE INFO

Article history:
Received 16 April 2015
Received in revised form 19 January 2016
Accepted 20 January 2016
Available online 29 January 2016

Keywords:
Visual language
Comics
Constituent structure
Visual cognition
Narrative
RSVP

ABSTRACT

Recent research has shown that comprehension of visual narrative relies on the ordering and timing of sequential images. Here we tested if rapidly presented 6-image long visual sequences could be understood as coherent narratives. Half of the sequences were correctly ordered and half had two of the four internal panels switched. Participants reported whether the sequence was correctly ordered and rated its coherence. Accuracy in detecting a switch increased when panels were presented for 1 s rather than 0.5 s. Doubling the duration of the first panel did not affect results. When two switched panels were further apart, order was discriminated more accurately and coherence ratings were low, revealing that a strong local adjacency effect influenced order and coherence judgments. Switched panels at constituent boundaries or within constituents were most disruptive to order discrimination, indicating that the preservation of constituent structure is critical to visual narrative grammar.

1. Introduction

Perceiving and integrating events over time is critical to the continuous dynamics of cognition (Corbetta & Shulman, 2002; Shipley & Zacks, 2008; Spivey & Dale, 2006). Humans and other animals can detect both static and dynamic changes in their environment over time (Hagmann & Cook, 2013; Rensink, 2004; Wright et al., 2010), but only humans can integrate information into a *narrative*, in which events depicted visually are interpreted as related and consequential to each other. Such comprehension is critical to understanding plots, stories, and instructions, and involves balancing a variety of covarying elements such as order, duration, and emphasis of component parts. While verbal narratives have been extensively researched, visual narratives have not, despite their prevalence in human culture for thousands of years—whether found on cave paintings, tapestries, or, in contemporary society, in the sequential images of comics (Kunzle, 1973; McCloud, 1994). Research into the comprehension of visual narratives has only recently emerged with seriousness and a focus on cognition (Cohn, 2013a; Magliano & Zacks, 2011; McCloud, 1994). We here explore one facet of this broader comprehension: the demands placed on perception and cognition in a rapidly presented visual narrative sequence.

Early research on sequential image understanding focused on the linear relations between images. Image-by-image comprehension involves continuously updating aspects of comprehension that rely on rapid scene understanding (Greene & Oliva, 2009; Potter, Wyble, Hagmann, & McCourt, 2014), and observing the changes that occur across characters, spatial location, and time (Magliano & Zacks, 2011; McCloud, 1994; Saraceni, 2001). Shifts in these dimensions (e.g., the introduction of a new character) incur costs in processing as a mental model of the narrative becomes updated with new information (Magliano, Dijkstra, & Zwaan, 1996; Magliano & Zacks, 2011; Zwaan & Radvansky, 1998).

Beyond these linear relations between images, *Visual Narrative Grammar* (VNG) argues that images in sequences take on narrative roles that are then combined into hierarchic constituents analogous to the way that sequential words take on syntactic roles that combine into constituents in sentences (Cohn, 2013b). This analogy is one at the functional level: a narrative grammar packages discourse-level meaning into a sequence using architectural constraints (categories, hierarchy, etc.) that are similar to the way that syntax packages meaning in sentences. VNG thus finds surface similarities with previous "grammatical" approaches to discourse (e.g., Clark, 1996; Hinds, 1976) particularly the well-known "story grammar" paradigms (e.g., Mandler & Johnson, 1977; Rumelhart, 1975), but differs from these precedents in both theoretical formalisms and the experimental methods used to provide evidence (see Cohn, 2013b for details). Experimentation has supported the idea that narrative structure in visual sequences is separate from its semantics (Cohn, Paczynski, Jackendoff, Holcomb, & Kuperberg, 2012), organized into constituents (Cohn, Jackendoff, Holcomb, & Kuperberg, 2014) and involves narrative categories defined by both content and context (Cohn, 2014).

* Corresponding author at: 426 Ostrom Ave, Syracuse, NY 13210, United States.
E-mail address: cehagman@syr.edu (C.E. Hagmann).

http://dx.doi.org/10.1016/j.actpsy.2016.01.011
0001-6918/

These structures can best be understood through an example. Fig. 1 illustrates a visual sequence with two constituents where Schroeder is playing in a sandbox with Snoopy. The sequence starts with an "Establisher" which sets up the situation of him playing in the sandbox. An "Initial" panel begins the events of the sequence, as Schroeder suddenly feels the heat of the sun. He then vigorously builds a sand mound in the subsequent "Peak" panel, a narrative climax of the primary actions of the sequence. The sequence is then resolved in the next panel, a "Release"—a resolution, aftermath, or coda of an action—where he rests in his newfound shade. A second constituent then begins suddenly with an even more climactic Peak, with Snoopy suddenly blowing the sand onto Schroeder, who then finds himself coated in the sequence-ending Release.

Importantly, narrative categories apply both to panels and to whole constituents. Together, the first four panels form their own constituent (an Initial) that, as a whole, sets up the entire second constituent (a Peak) at a higher level of structure. Each constituent is motivated internally by a Peak, which acts as the "head" of that constituent (double-barred lines). The penultimate panel of Snoopy blowing sand is thus the narrative climax of the whole sequence, reflected in its status as the Peak panel motivating the Peak constituent. The canonical *Establisher–Initial–Peak–Release* pattern is thus used in part or full at various levels of structure. These top–down global structures interact with the bottom-up content of images to determine the roles that images play in the sequence (Cohn, 2013b, 2014).

Initial evidence for the psychological reality of this narrative grammar came from experiments that used 1500 ms/panel sequences that balanced the contributions of narrative structure and/or semantic associative relationships across images (Cohn et al., 2012). Response times to panels in a target monitoring task were faster for panels in normal sequences, with both structure and meaning, than fully scrambled sequences of images with no narrative and no meaningful relations across images. However, intermediate response times resulted from target images in sequences with only semantic associations and no narrative structure (i.e., that maintained a common theme across panels) and from sequences with only narrative structure but no semantic associations (i.e., visual narrative analogs to a sentence like *Colorless green ideas sleep furiously*, which has syntax but no clear meaning). Across all sequences types, response times decreased across the ordinal position of sequences. Such results showed that narrative structure provides a behavioral advantage to the processing of sequences.

Another experiment using the same stimuli measured event-related brain potentials (Cohn et al., 2012), specifically the N400, a neural response typically lasting from 250 to 500 ms peaking around 400 ms, and thought to reflect the activation state of an incoming stimulus in semantic memory (Kutas & Federmeier, 2011). Panels from sequences with only semantic associations produced larger amplitude N400s than normal sequences. Even larger N400s appeared with scrambled and narrative-only sequences. Both scrambled and narrative-only sequences lacked coherent semantic associations between images, but the narrative-only sequences did have a felicitous narrative structure. Yet, because these amplitudes did not differ between scrambled sequences and narrative-only sequences, it confirmed that this narrative grammar was different from meaning, since the N400 was not attenuated by the presence of narrative structure. The N400 was, however, attenuated across ordinal panel position only in normal sequences, suggesting that a facilitation of meaning only occurs in the presence of both coherent narrative and semantic associations across images. Thus, while the low-level semantic information between images is involved in the comprehension of sequential images, it interacts with the global narrative structure.

Other studies have used techniques of rearranging images in visual narrative sequences to analyze their global structure and the roles taken by panels within a sequence. When viewing sequences at their own pace, comprehenders spend more time viewing panels from fully scrambled sequences than from coherent sequences (Cohn & Wittenberg, 2015; Foulsham, Wybrow & Cohn, submitted for publication). This slowing even occurs on the opening image of a sequence, where context has not yet rendered a sequence as incomprehensible, which suggests that some images are better candidates to open a narrative sequence than others (Cohn & Wittenberg, 2015; Cohn, 2014; Foulsham et al., submitted for publication). More targeted switching of panel positions within 4-panel sequences showed that comprehension worsens when panels are switched across distances than when switched locally (Cohn, 2014). Similarly, when participants were given four panels and asked to arrange them in an order that makes sense, misplaced panels were moved to adjacent positions more often than positions further in a sequence (Cohn, 2014). This *adjacency effect* was likely related to some images being more central to the narrative, and being surrounded by more peripheral images, which play more flexible roles in the sequence.

This global scope of narrative structure also must take into account the constituents formed by groupings of panels. Studies have long shown that participants are highly consistent in where they choose to divide both drawn and filmed visual sequences into sub-episodes (Cohn & Bender, submitted for publication; Gernsbacher, 1985; Magliano & Zacks, 2011). While research on this segmentation has typically viewed changes in linear coherence (such as shifts in characters or locations) as indicative of constituency boundaries (Gernsbacher, 1985, 1990; Magliano & Zacks, 2011; Radvansky & Zacks, 2014; Zacks, 2014), research within the VNG paradigm has shown that constituent structures go beyond such transient semantic changes. For example, narrative category information has been shown to be more predictive of conscious segmentation of drawn visual sequences than linear coherence relationships, though both do influence such divisions (Cohn & Bender, submitted for publication). Furthermore, measuring event-related potentials, Cohn et al. (2014) found that blank "disruption panels" placed within the constituents of visual narratives elicited a larger left anterior negativity than those placed between constituents, and this neural response was similar to those evoked by manipulations of grammar in language (Hagoort, 2003; Neville, Nicol, Barss, Forster, & Garrett, 1991). This effect could not be attributed to changes in linear coherence, because the amplitude to disruption panels was greater for

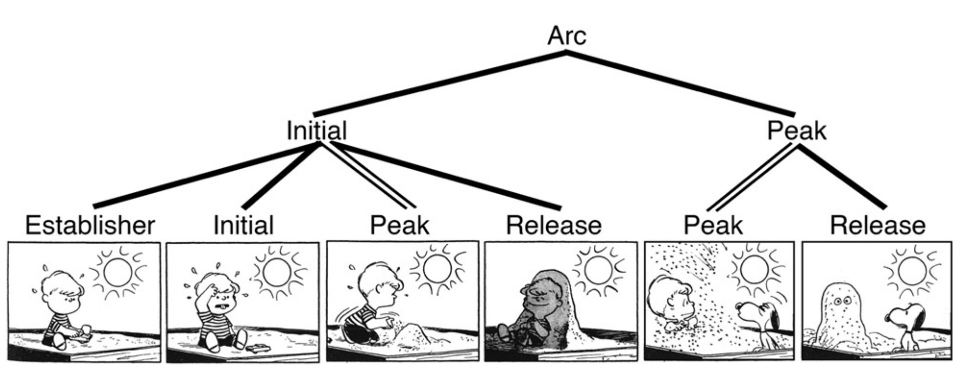

Fig. 1. Structure of a visual sequence with narrative categories and constituents.