COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 38 (2014) 1317–1359

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12116

You're a Good Structure, Charlie Brown: The Distribution of Narrative Categories in Comic Strips

Neil Cohn

Center for Research in Language, University of California, San Diego

Received 18 September 2012; received in revised form 26 April 2013; accepted 3 July 2013

Abstract

Cohn's (2013) theory of "Visual Narrative Grammar" argues that sequential images take on categorical roles in a narrative structure, which organizes them into hierarchic constituents analogous to the organization of syntactic categories in sentences. This theory proposes that narrative categories, like syntactic categories, can be identified through diagnostic tests that reveal tendencies for their distribution throughout a sequence. This paper describes four experiments testing these diagnostics to provide support for the validity of these narrative categories. In Experiment 1, participants reconstructed unordered panels of a comic strip into an order that makes sense. Experiment 2 measured viewing times to panels in sequences where the order of panels was reversed. In Experiment 3, participants again reconstructed strips but also deleted a panel from the sequence. Finally, in Experiment 4 participants identified where a panel had been deleted from a comic strip and rated that strip's coherence. Overall, categories had consistent distributional tendencies within experiments and complementary tendencies across experiments. These results point toward an interaction between categorical roles and a global narrative structure.

Keywords: Narrative structure; Discourse; Comics; Film; Visual language

1. Introduction

From Aristotle's exploration of theatric plots (Butcher, 1902) to psycholinguistic theories of "story grammars" (e.g., Mandler & Johnson, 1977; Rumelhart, 1975; Thorndyke, 1977), the structure of narrative has been recognized as a central part of human expression. While most research on the psychology of narrative has focused on verbal discourse, visual narratives of sequential images have persisted for millennia, from cave paintings to

Correspondence should be sent to Neil Cohn, Center for Research in Language, University of California, San Diego, 9500 Gilman Dr., Dept 0526, La Jolla, CA 92093-0526. E-mail: neilcohn@visuallanguagelab.com

modern comics (Kunzle, 1973). In a recent theory, Cohn (2003, 2013) argues that a "Visual Narrative Grammar" (VNG) guides the comprehension of sequential images analogously to syntactic structure in sentences. Like words in sentences, this grammar organizes narrative categories into hierarchic constituents. However, because "a picture is worth a thousand words," this grammar operates at a level above that of individual sentences, closer to that found in discourse and narrative.

How can we explore the existence of the categories proposed by this theory? Linguistics research uses various diagnostics to determine syntactic categories, such as moving elements around or deleting them, and similar techniques can be used in visual narratives (Cohn, 2013). This paper describes experiments based on these diagnostics for narrative categories outlined in VNG which asked participants to order randomly arranged panels of a comic strip, to delete panels of their choice, and to recognize from where in the strip panels were deleted. Together, these tasks provide converging evidence that the roles of narrative units—at least in sequential images—are influenced by images' content as well as their context in the broader sequence.

1.1. The structure of visual narrative

In VNG, individual panels take on narrative categorical roles, which are structured into constituents. This theory extends beyond focusing on the changes between linear relationships between adjacent discourse units, be they static or moving images (Magliano & Zacks, 2011; McCloud, 1993; Saraceni, 2003; Zwaan & Radvansky, 1998) or text, such as adjacent sentences or clauses (Halliday & Hasan, 1976; Hobbs, 1985; Kehler, 2002; Mann & Thompson, 1987; Trabasso, Secco, & van den Broek, 1984). It also differs from previous grammatical approaches to narrative such as the story grammars of the late 1970s, which focused on characters' goal-directed actions (e.g., Mandler & Johnson, 1977; Rumelhart, 1975; Stein & Glenn, 1979; Thorndyke, 1977). Insights of these previous approaches have been integrated into this narrative grammar (for example, linear shifts may signal breaks between constituents), although ultimately important differences exist between these models, discussed in depth in Cohn (2013).

Narrative categories in VNG are assigned through correspondences with prototypical semantic features, as well as through their context in a global narrative structure. A sequence may begin with an Establisher, which sets up the situation without acting on it. An Initial then begins the events of the sequence, which may be extended throughout a Prolongation. The narrative then reaches its climax in the Peak, which is followed by a Release that resolves the sequence. This arc is operationalized into a single "phase" structure rule: *Phase X→ (Establisher) – (Initial (Prolongation)) – Peak – (Release)*

This rule states that narrative categories progress in a canonical order, although only the Peak is fully necessary (parentheses indicate optionality of all other categories). Not only do narrative categories apply to individual panels but they also apply to sequences of panels that themselves form a phase. This means that groupings of panels can play narrative roles at a higher level of the grammar. An "arc" for narrative is then analogous to a "sentence": It is a maximal node with no other role in the narrative.

Consider Fig. 1. This sequence shows Charlie Brown pitching a baseball, only to have it hit right back at him. The strip begins with Charlie reaching back to throw the ball, a prototypical Initial because it shows a preparatory action. This action is completed in a Peak in the next panel when he throws the ball. The next panel is another Peak—it shows the culmination of the strip as the ball is hit back at Charlie. The final panel is a Release, showing his consternation as the aftermath of the event. Together, the first two panels set up the culmination in the third panel. These panels form an Initial at a higher node, motivated by the source of a path in the subordinate Peak (depicted by double-bar lines). Thus, the second panel plays two roles—it is locally the Peak for the first panel, and it motivates a higher level Initial for the subsequent sequence. In addition, while the linear sequence of the strip does not conform to the canonical pattern (*Initial-Peak-Peak-Release*), its constituents do: first with a short phase of *Initial-Peak* at a local level and second with an *Initial-Peak-Release* at the higher level.

Importantly, this model keeps narrative structure and semantics separate (Cohn, Paczynski, Jackendoff, Holcomb, & Kuperberg, 2012). Certain semantic meanings prototypically correspond to particular categories, but non-prototypical mappings also occur. In such cases, a panel's context in the sequence may influence its category more than its content. This is similar to grammatical classes in sentences, which prototypically map to certain semantic features but ultimately are assigned through distributional patterns in a sentence (Jackendoff, 1990). For instance, contrary to notions that nouns are "people, places, things, and ideas," words become nouns because of their behavior within a sentence. Nouns can be pluralized, and in English often follow determiners or adjectives, follow prepositions, and so on. Comparable patterning applies to other grammatical classes, and by analogy, to narrative categories.

In order to better frame the correspondences between narrative and meaning, we discuss properties of each category in further detail below. For a more comprehensive description of these properties and the architecture of narrative grammar, see Cohn (2013).

Peaks (P) form the core of the narrative arc. They show the most important aspects of the arc, such as the culmination of events, the apex of the narrative, or the goal in a

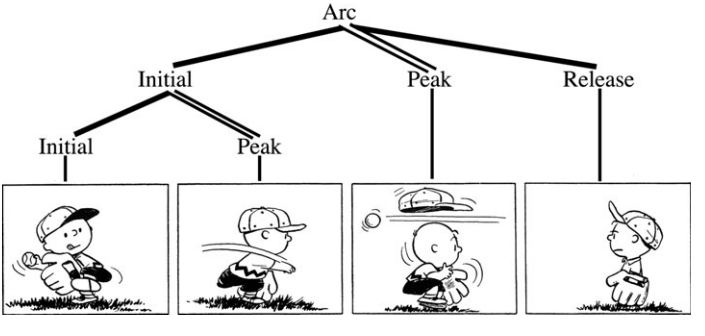

Fig. 1. Narrative constituent structure in a four-panel sequence (Schulz, 2004a). Peanuts is © Peanuts Worldwide LLC.

trajectory of a path. For example, in Fig. 1 panel 2 is a Peak because it shows a completed action (throwing the ball), while panel 3 is a Peak because it is a culminating action (the ball almost hits Charlie). However, the Peak in the third panel of Fig. 2a is not the culmination of an event that begins in a prior panel—it is an interruption of those events. Yet it is still the climax of the sequence, and thus a Peak. Because Peaks feature such key events, they show the result of causal actions.

Initials (I) are the second most important category, since they lead to the Peak. These prototypically show a preparatory action (as in panel 1 of Fig. 1), the inception of an event, and/or causal actions. They may also show the source of a trajectory along a path toward a goal (as in panel 1 of Fig. 1). But again, non-prototypical correspondences from semantics to this category are also possible. Fig. 2a has an Initial in the second panel where Snoopy is chasing the hockey puck. It does not show him beginning this action or preparing to run—it jumps directly into the process of chasing. Thus here, the relative context to the Peak determines it as an Initial more than its depicted actions.

Consider also the second panel in Fig. 1. It shows the culmination of an event like a Peak (throwing a ball) but also the source of a path like an Initial (the ball's trajectory). In this case, the completed action resolves the event motivated by the preceding Initial (the throwing), and thus is a Peak locally. However, this completed action causes the source of the path, thereby motivating this Peak panel to be the "head"—the motivating panel—of a larger phase which acts as an Initial at a higher level of structure. In this way, narrative categories apply recursively both to narrative units (panels) and narrative

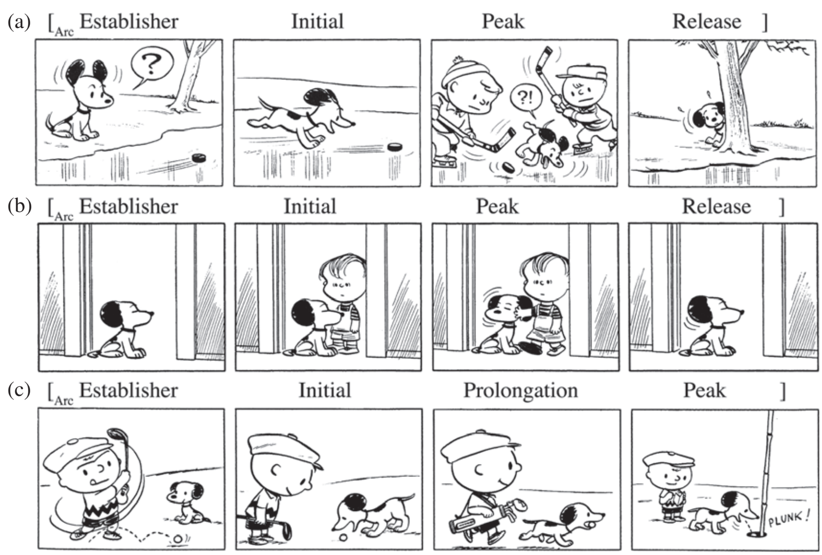

Fig. 2. Visual narrative sequences (Schulz, 2004a): (a) a canonical narrative arc with a non-prototypical mapping of narrative and events, (b) narrative sequence with the same panel acting as Establisher and Release, (c) a strip with a Prolongation panel extending the scene between Initial and Peak. Peanuts is © Peanuts Worldwide LLC.

constituents (groups of panels). Panels with these dual functions as local Peaks "downstairs" in a constituent and heads of Initials "upstairs" for a whole phase will be called *Initiating Peaks*.[1]

Releases (R) often follow Peaks. These panels prototypically show an aftermath of events or the coda of an action, often as passive states relative to actual actions. For example, the Release in the final panel of Fig. 1 shows Charlie's response to the hit baseball. Fig. 2b uses a Release in the final panel to return Snoopy to the same passive state he began in (discussed further below). In contrast, the Release in Fig. 2a shows Snoopy's reaction to the hockey players. Here, Snoopy hiding is a whole new action, not the aftermath of the events in the Peak (such a Release might show him flattened on the ice). While these Releases have different semantic characteristics, *narratively* they all provide resolutions to preceding events.

Establishers (E) may open a sequence or constituent by setting up the characters and context of the interaction. Prototypically, this is a passive state, as in the first panel of Fig. 2b, where Snoopy does not engage in any action. Here, referential information is provided that may be acted upon later. This category is similar to "establishing shots" from film (Arijon, 1976; Bordwell & Thompson, 1997) and comics creation (Abel & Madden, 2008; McCloud, 2006), and possibly with discourse notions of Topics (e.g., Clark, 1996) or "setting the stage" (Hinds, 1976). Unlike these notions though, Establishers are not restricted to introducing characters for a broader discourse, but rather establish referential information constrained to the subsequent sequence. Thus, Establishers can appear in subconstituents of arcs where characters have already been introduced, though they may engage in new actions or contexts. Establishers also do not always depict passive states, as in the first panel of Fig. 2a where both characters are engaged in actions: The puck is in the process of moving and Snoopy is surprised by it. However, this panel *establishes* the relationship between these characters before launching into their actions.

Establishers may also share a relationship with Releases. In Fig. 2b, the first and last panels show the same *semantic* features (Snoopy sitting in a doorway with head straight forward), yet they play different *narrative* roles based on their position in the sequence. Starting and ending a narrative in the same place is common among narrative theories, such as "Freytag's triangle" (1894), where the exposition and resolution are the same. Events also share this structure, possibly beginning and ending in the same physical place (like jumping; Lasher, 1981), or generally start and end in a passive state (Talmy, 2000). Thus, although Establishers and Releases may depict the same prototypical semantic properties—passive actions—they play different functional roles in the narrative.

Prolongations (L) are situated between the Initial and the Peak. They often extend the actions of the Initial, act as a buffer or rhythmic pause between Initial and Peak, or elaborate on the trajectory or manner of a path. Take for example Fig. 2c, which starts with an Establisher of Charlie hitting a golf ball, which establishes the relationship of Snoopy and the ball. Snoopy picks up the ball in the Initial and puts it into the cup at a Peak, the final panel. The penultimate panel just extends the action of the first, providing a medial state in the progression toward the cup, while building narrative tension before the Peak. This panel is a Prolongation because of this elaborative role.

Altogether, these categories form a narrative sequence. Research in the story grammar tradition has established that comprehenders prefer a canonical narrative arc to other patterns. Participants' remembered stories better if they conformed to the canonical story structure than if stories manipulated the temporal order of the events described (Mandler & Johnson, 1977), inverted the order of sentences (Mandler, 1978, 1984; Mandler & DeForest, 1979), or scrambled the orders of sentences (Mandler, 1984). Overall, recall worsened based on how much a story departed from the canonical pattern (Stein & Nezworski, 1978). Additionally, recall tended to focus on higher level organizational aspects of the narrative rather than content-specific details (Thorndyke, 1977). These findings suggested that a global canonical narrative pattern stored in long-term memory aids comprehension.

Previous research has provided evidence of the psychological validity of the theory of Visual Narrative Grammar (Cohn et al., 2012). We presented participants with four types of sequences. Normal sequences had both narrative structure and semantic associations between panels. Other sequences had no overarching narrative arc but did have semantic associations between panels (various panels showing scenes about baseball but no coherent narrative). Still other sequences had no semantic associations between panels but had a coherent narrative structure (analogous to sentences like *Colorless green ideas sleep furiously* which have syntax but no semantics). Finally, some sequences scrambled panels with neither semantic associations nor narrative structure.

We found that participants were fastest to monitoring target panels in the normal sequences, were slowest to panels in sequences with scrambled panels, and had intermediate times for panels in sequences with only semantic associations or narrative structure. This indicated that narrative structure gives an advantage in processing sequential images, although not as much as with semantic associations between panels. An additional experiment using event-related potentials showed that the N400 effect—typically associated with semantic processing (Kutas & Federmeier, 2011; Kutas & Hillyard, 1980)—did not differentiate between panels from scrambled sequences and those with only narrative structure. This suggested that, although narrative structure gives a behavioral advantage in processing, this structure is different from semantics (panels from sequences with semantic associations did show an attenuated effect).

Together, these studies provided initial support for the existence of a narrative structure in sequential image processing as being separate from semantic associations. However, they did not provide evidence for the unique roles of narrative categories in a sequence. For example, are Establishers and Prolongations more expendable because they show referential information that is available elsewhere in a sequence? What relationship do Establishers and Releases have as categories? Are Initials and Peaks really the focal panels of a sequence? Do categories behave differently in canonical versus non-canonical sequences?

Searching for evidence of narrative categories can draw upon the analogy to syntactic categories in sentences. As discussed, grammatical classes like nouns and verbs typically correspond to semantic objects and events, although ultimately these categories are determined by their distribution throughout a sentence (Jackendoff, 1990). Linguistics has developed various diagnostics to test for these categorical roles, such as moving or deleting words in a sentence. Extending this analogy to narrative, Cohn (2013) argues that moving or deleting narrative categories can also reveal different distributional tendencies.

We therefore designed four experiments that sought converging evidence for these distributional regularities using complementary tasks. Experiment 1 used a "reconstruction task" asking participants to arrange four unordered panels in a coherent sequence. This task investigated whether certain categories tend to be misplaced more than others, or if panels originally appearing in one position might commonly be assigned to another. Some of these results were followed up in Experiment 2 by measuring how long participants viewed each panel on a screen as they progressed through sequences where the order of panels had been reversed. Experiment 3 repeated the reconstruction task from Experiment 1, except participants also chose one panel to delete. This experiment tested how important or expendable certain categories might be to a sequence. Finally, Experiment 4 probed the reverse intuitions by presenting three-panel strips and asking what position a panel was deleted from, and asking participants to rate the coherence of the strip.

If categories have consistent distributional regularities across these tasks, it would support categorical roles in a narrative. Because they contain information pertinent to the core events of the sequence, Initials and Peaks should be misplaced less and should not be easily deleted. In contrast, Establishers, Releases, and Prolongations should be more expendable, misplaced in complementary positions, and able to be deleted with little recourse. Such findings would contrast with theories claiming that any discourse unit can take a different meaning depending on its context in a narrative (Jahn, 1997; Sternberg, 1982). A strong version of this principle argues that *any* surface form could appear at *any* place in a narrative and still be meaningful. As stated, VNG predicts that, while certain panels possibly play various roles, this is determined by a combination of their semantic content and the context imposed by the sequence. Thus, some panels should be more flexible in playing various roles than others, based on their content.

2. Experiment 1: Reconstruction

Our first task asked participants to order unarranged panels into a sequence that made sense. This technique has been commonly used to research the comprehension of sequential images (Bresman, 2004; Huber & Gleber, 1982; Lynch & Hagen, unpublished data; Nakazawa & Nakazawa, 1993) and is used as a measure of "logical/sequential reasoning" in the WAIS-III test of non-verbal IQ (Kaufman & Lichtenberger, 2006). In general, people are highly proficient at arranging strips into their original orders, although this ability increases with age and experience in reading comics (Nakazawa, 2004; Nakazawa & Nakazawa, 1993).

Some unpublished work has also shown regularities among types of panels within a sequence. Panels chosen to start strips often "set the stage," begin an event, and feature a wide "long shot" viewpoint (Lynch & Hagen, unpublished data), while panels chosen to end strips often show the completion or coda of an event, or the punchline of a joke (Bresman, 2004; Lynch & Hagen, unpublished data). In general, panels containing fewer events are misplaced further from their original placement than those with content more relevant to the central message of the strip (Bresman, 2004). These studies show that reconstruction tasks can provide evidence about panels' roles in relation to a narrative sequence.

Experiment 1 thus used a similar reconstruction task to seek evidence for the categories described in the theory of narrative structure described above, using strips with canonical and non-canonical patterns. We first asked how accurate participants would be in reconstructing the whole strip. If participants rely only on a canonical narrative structure (e.g., Bordwell, 2007; Bordwell & Thompson, 1997; Mandler & Johnson, 1977; Rumelhart, 1975; Stein & Glenn, 1979; Thorndyke, 1977), non-canonical sequences should be harder to reconstruct than canonical ones. However, if both types are reconstructed at the same rates, participants may be using other strategies, like linear semantic panel-to-panel relationships or, as hypothesized in VNG, relying on an interaction between the semantic content of individual panels and their context in the sequence. To further assess these possibilities, we also looked at the rates of misplacement for original positions and narrative categories.

We next asked how accurate participants would be in placing panels in their original positions. If narrative comprehension uses only linear semantic relationships (e.g., Magliano & Zacks, 2011; McCloud, 1993; Zwaan & Radvansky, 1998), all positions of panels should be misplaced with the same frequency or be distributed in a sequence at a chance probability. This would occur because any two panels can create *some* type of linear semantic relationship, even if it is non-sensical (McCloud, 1993; Saraceni, 2000, 2003), and thus all panels would have an equal likelihood of following another. If a global canonical pattern alone determines the role of a narrative unit, misplacement should be dictated by sequence *position* but not by the distributional tendencies of categories. Here, units play roles in the global context, but not motivated by their content. Such a perspective could be dispelled if narrative categories have the same tendencies in both canonical and non-canonical sequences. This latter view would be predicted by VNG, since a canonical narrative structure does aid comprehension, although alternative sequences are possible.

Next, we looked at the specific locations that misplaced categories were placed and at the full strip patterns that resulted from misplaced panels. Again, if only linear semantic relationships guide comprehension, all positions of panels should be misplaced with the same frequency or be distributed in a sequence at a chance probability. If a global canonical pattern determines the role of a narrative unit, categories should have no consistent placement throughout a sequence and categories in canonical sequences should be misplaced at lower rates than those from non-canonical sequences. However, if categories become misplaced at the same rates in canonical and non-canonical sequences, it would support that categories play roles in a sequence beyond the global canonical pattern.

Finally, VNG predicts that narrative categories balance semantic properties and global context. Thus, individual categories should have unique distributional tendencies that persist across canonical and non-canonical sequences while also showing effects of canonical sequence position. According to VNG, Initials and Peaks are fairly stable in their placements while other less essential categories like Establishers and Releases may fall in complementary distribution. For example, Fig. 3a depicts a Release that has been moved to the front of the strip, while Fig. 3b shows an Establisher moved to the end. While the meanings of the strips change slightly, the sequences remain felicitous *as a narrative*. In contrast, displacement of the Peak to the front (3c) or the Initial to the end (3d) creates less coherent sequences. Thus, Initials and Peaks should be misplaced in a sequence at low rates, while Establishers and Releases (and maybe Prolongations) will move frequently into each other's positions. Such findings would support that narrative categories are influenced by both their content and their context.

2.1. Methods

2.1.1. Stimuli

We culled 180 black and white comic strips from the first two volumes of the *Complete Peanuts* (Schulz, 2004a, b) by Charles Schulz. *Peanuts* was chosen because (a) it has a large archive, (b) its strips are short with preestablished boundaries, (c) all of its strips maintain a four-panel structure with consistent panel sizes, (d) many strips are without text, and (e) the strips are clear in their representations of events and narrative structures. Only strips without text were used, or text was deleted from panels beyond non-verbal signs (like question marks or exclamation points), text within the fictive world of the strip (such as letters on building blocks or street signs), and some sound effects that clarified the visual actions.

The 180 strips were randomly divided into six lists, which distributed the strips in a counterbalanced fashion throughout the tasks of Experiments 1, 3, or 4. Though a

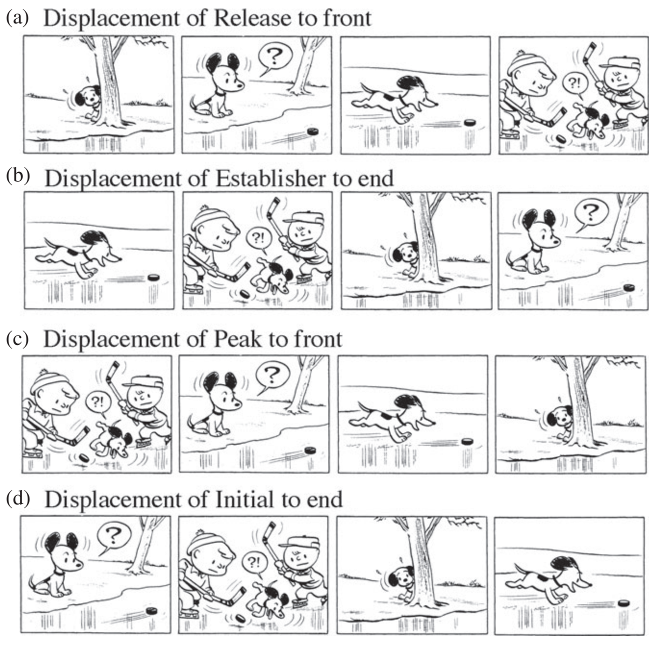

Fig. 3. Movement of narrative categories and their effect on structure.

participant experienced all three experimental tasks in a single session, each strip was viewed only once by each participant and no strip was repeated across tasks (i.e., each task used only 60 strips, with all 180 strips viewed in total across the three tasks). Strips were randomly ordered within the blocks of a task in each list. The blocks of experimental tasks were also counterbalanced into three different "task orders," such that each participant experienced all tasks in different orders depending on the list but had to complete a whole section before moving on to the next task (Task order 1: Exp. 1, 3, 4; Task order 2: Exp. 4, 1, 3; Task order 3: Exp. 3, 4, 1).

Prior to experimentation, strips and panels were coded for narrative categories using the semantic criteria described above and other diagnostics outlined in Cohn (2013). Nearly half of all strips (88) featured a narrative in the complete canonical *Establisher-Initial-Peak-Release* ("EIPR") pattern, while the remaining 92 featured a diverse variety of other patterns (i.e., IPIP, ELPR, IIIP, etc.). Frequencies of narrative categories and their proportion of total panels are summarized in Table 1 for EIPR and Non-EIPR strips. These percentages were used to determine "chance" in the experiments described below.

The coding of panels in strips reinforced that, motivated by the canonical pattern, certain types of categories gravitate toward particular locations in a strip. Establishers naturally were at the front, Releases appeared toward the end, and Prolongations always fell in the middle. Initials and Peaks, however, appeared more flexible in their distribution. Relative positioning further constrained these categories: Establishers and Initials preceded corresponding Peaks, while Releases followed them, and Prolongations fell between Initials and Peaks. Table 1 summarizes the distribution of categories across all four positions in the original strips: Establishers appear only in the first two positions, Initials never end a strip, Prolongations never begin or end a strip, and Releases never start a strip.

2.1.2. Participants

Comic readers from Tufts University (10 male, 2 female, mean age: 19) were paid for their participation. Each participant gave informed written consent according to the guidelines of the Tufts University Institutional Review Board. All participants took a pretest

Table 1
Frequency and percentage of narrative categories in strip patterns, as well as distribution of narrative categories across panel position

	Qty/%						
	All Strips	EIPR	Non-Canonical	Position 1	Position 2	Position 3	Position 4
Establisher	141/.20	88/.25	53/.14	130/.72	11/.06	–	–
Initial	199/.28	88/.25	111/.30	29/.16	123/.68	47/.26	–
Prolongation	15/.02	–	15/.04	–	5/.03	10/.06	–
Initiating Peak	41/.06	–	41/.11	17/.09	21/.12	3/.02	–
Peak	185/.26	88/.25	97/.26	4/.02	8/.04	111/.62	62/.34
Release	189/.19	88/.25	51/.13	–	12/.07	9/.05	118/.66
Total strips	180	88	92				

questionnaire that rated their current and childhood comic reading habits (comic strips, comic books, Japanese comics, and graphic novels). These ratings were measured using a scale of 1 to 7 (1 = *never*, 7 = *always*), and the questionnaire also gauged their self-assessed "expertise" at reading and drawing comics along a five-point scale (1 = *below average*, 5 = *above average*). A "fluency rating" was then computed using the following formula:

$$(\text{Mean Comic Reading Freq} \times \text{Comic Reading Expertise}) + \left(\frac{\text{Comic Drawing Freq} \times \text{Drawing Ability}}{2}\right)$$

This formula weighted fluency toward comic reading comprehension, giving an additional "bonus" for fluency in comic production (Cohn et al., 2012). Participants' fluency was above the idealized average (a score of 12) to high (22), with an average score of 19.09 ($SD = 9.5$). Studies have indicated differences in comprehension for experienced versus novice comic readers (Cohn et al., 2012; Nakazawa, 2002), and thus by using comic "fluent" participants, we aimed to reduce the heterogeneity of the population. Finally, all participants indicated that they knew *Peanuts* specifically, but posttest questionnaires revealed that they did not recognize the specific strips used in the experiments.

2.1.3. Procedure

In an open PowerPoint file, participants viewed four panels in a cross (+) formation and were asked to arrange them into a linear order that "makes sense." No emphasis was placed on creating the "correct" order—just one that satisfied the participant's comprehension. The starting position of panels was randomized to reduce the likelihood that participants would use a strategy to reconstruct the strips. Only panels from one strip were shown at a time and participants controlled the pace at which they moved to the next strip. Most found the task relatively easy and enjoyable.

2.1.4. Data analysis

We analyzed reconstructed strips by coding panels' final positions relative to their original position. For example, if a canonical strip's panels 1 and 2 were misplaced, panel 1 was coded as "1:2" and 2 was coded as "2:1," meaning that the Establisher was misplaced into second position, while the Initial was misplaced into first position. Meanwhile, if panels 3 and 4 then were accurately positioned, they were coded as 3:3 and 4:4. From these codings, we analyzed participants' average accurate and inaccurate reconstructions of individual categories, panel positions, and full strip patterns.

Misplaced panels were compared in a subjects' analysis collapsing across stimuli items for positions and narrative categories using repeated-measures Analysis of Variance (ANOVA). Follow-up analyses used independent t tests to compare rates of misplacement against chance. Chance for any full strip pattern was calculated as the 1 in 24 possibility of any ordering of a four-panel strip (4 positions × 3 positions × 2 positions ×

1 position = 24, 1/24 = .04). Chance for individual positions was calculated as the 1 in 16 possibility of any position being displaced to any other position (4 positions $\times$ 4 positions = 16, 1/16 = .0625). Chance for categories was calculated as the percentage that a category appears in all strips (summarized in Table 1). Additional Pearson correlations compared misplaced categories with each other, with positions, and with inaccurate orderings of full strips.

Additionally, one-way ANOVAS for each position/category under analysis also introduced "task order" as a between-subjects factor. This statistic examined whether the order in which the Reconstruction Task appeared relative to other tasks from Experiments 3 and 4 influenced the results. These analyses were initially carried out for the combined results (EIPR and Non-EIPR), with follow-up analyses performed for significant findings at individual strip patterns.

2.2. Results

Overall, 77% of whole strips were accurately reconstructed, while 83% of individual panels retained their original placement. Participants were also more accurate for reconstructing EIPR strips (81%) than Non-EIPR strips (73%), a near significant difference, $t(11) = 2.03$, $p = .067$.

2.2.1. Positions

Rates of accurate placement did differ between positions, $F(3, 33) = 4.84$, $p < .01$. No significant main effects were found of task order for any position (all Fs < .298, all ps > .749). Next, we asked: Which panel positions were placed in their original positions more than others? Panels originally from Position 2 ($M = .80$, $SD = .08$) were misplaced more than those from Position 1 ($M = .84$, $SD = .06$), and Position 4 ($M = .85$, $SD = .09$), while panels from Position 4 were misplaced significantly more than Position 3 ($M = .81$, $SD = .09$; all ts > 3.0, all ps < .05). Finally, we asked whether these rates of accurate reconstruction were greater than chance, which was indeed the case for all positions (all ts > -3, all ps < .01).

2.2.2. Narrative categories

Overall, accuracy for reconstructing narrative categories was consistent (summarized in Table 2), but no significant differences were found between the rates of misplacement between all five narrative categories, $F(5, 55) = .445$, $p = .815$. Also, no main effects appeared for task order for any category (all Fs < 1.05, all ps > .389). Although they did not differ from each other, some rates of misplacement differed for categories compared to chance. Prolongations and Initiating Peaks were misplaced more than chance (all ts < -3, all ps < .01), and Initials and Peaks less than chance (all ts > 4, all ps < .01). Establishers and Releases were misplaced near chance.

Across all narrative categories, Non-EIPR strips had more misplacements than panels from EIPR strips, summarized in Table 2. However, rates of misplacement between individual categories from EIPR and Non-EIPR strips did not significantly differ. Again,

Table 2
Rates of misplaced narrative categories of different strip types

	Mean (*SD*)		
	All Strips	EIPR Strips	Non-Canonical Strips
Establisher	.17 (.09)	.14 (.10)**	.21 (.13)***
Initial	.19 (.07)*	.15 (.09)**	.20 (.09)*
Prolongation	.22 (.30)**	–	.25 (.31)***
Initiating Peak	.21 (.16)**	–	.21 (.16)***
Peak	.15 (.09)*	.13 (.08)**	.16 (.14)*
Release	.17 (.10)	.12 (.09)**	.25 (.19)***

Notes. *Significantly below chance; **significantly above chance; ***trending above chance.

however, some categories were misplaced at rates significantly different than chance. In EIPR strips, all categories were misplaced significantly higher than chance (chance = .0625, all ts > 2.2, all ps < .05). In contrast, in Non-EIPR strips, Initials and Peaks fell below chance, $t(11) = 3.72$, $p < .005$ and $t(11) = 2.45$, $p < .05$ respectively, while trends toward significance above chance were shown for Establishers, $t(11) = -1.82$, $p = .096$, Prolongations, $t(11) = -2.2$, $p = .055$, Initiating Peaks, $t(11) = -2.2$, $p = .05$, and Releases, $t(11) = -1.9$, $p = .072$.

2.2.3. Positions placed by categories

Our next question asked: what positions in a sequence were panels moved to? Overall, misplaced panels were put into non-canonical positions (i.e., *not* E-I-P-R to 1-2-3-4) significantly more often ($M = .08$, $SD = .03$) than to positions matching canonical orders ($M = .05$, $SD = .03$), $t(11) = -4.082$, $p < .01$. Also, panels were misplaced to adjacent positions ($M = .089$, $SD = .06$) more than to locations at least one panel away from their original position ($M = .057$, $SD = .03$), a difference trending toward significance, $t(11) = 2.16$, $p = .054$.

Fig. 4 summarizes the rates of misplacement for narrative categories into different positions. Misplaced categories were compared against the 25% chance that they could be placed in any particular position. Establishers were placed in Position 2 above chance, $t(11) = 3.64$, $p < .005$, and then Position 4 (not significant), far more than being misplaced in Positions 1 or 3, which were below chance, $t(11) = -4.93$, $p < .001$ and $t(11) = -4.6$, $p < .005$, respectively. Initials and Prolongations were generally fronted, with both moved most often to Position 1. However, only Prolongations misplaced into Position 3 were significant, and below chance, $t(5) = -3.5$, $p < .05$. Initiating Peaks were consistently misplaced into Position 3—the canonical Peak position—six times more than to any other position, $t(11) = 3.80$, $p < .005$, while movement to Position 1 was significantly below chance, $t(11) = -7.25$, $p < .001$. Finally, Peaks were misplaced the most into Position 2, and Releases were misplaced into Position 3 almost twice as much as any other location.

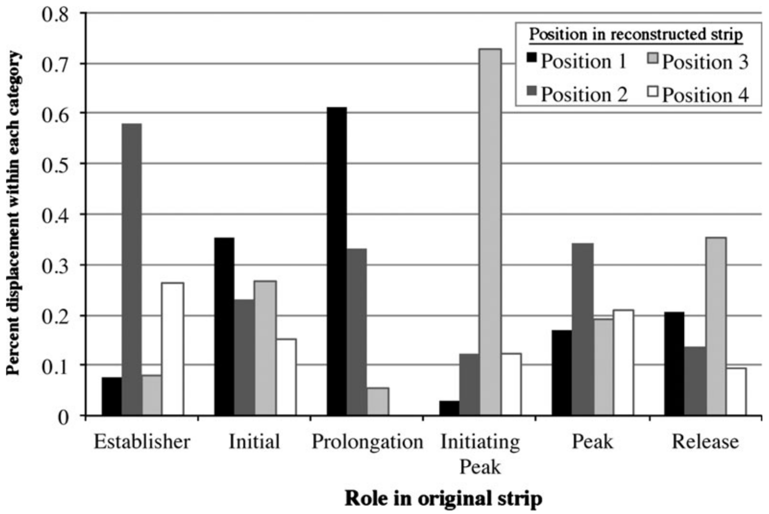

Fig. 4. Rates of misplacement into various positions by narrative categories.

2.2.4. *Categories within full strips*

Finally, the patterns of misordered full strips were analyzed to gain further insight on where categories were misplaced within a sequence. With 24 possible four-panel strips, each order had a probability of .04. All misorderings were treated the same, with no weight given to strips that preserved part of, but not the full, order (i.e., 4123 maintains the original order except for the fronting of 4). Against chance, no pattern other than the original appeared to a significant degree. Nevertheless, certain reconstructed patterns appeared at higher proportions than others and these patterns are worth discussing.

Only four patterns had a rate higher than .02 across all strip types *and* EIPR strips *and* Non-EIPR strips, shown in Fig. 5. We reasoned that correlations between the rates of reconstruction for these full strips and those of individual narrative categories might lend additional insight into which categories are moving into which positions. For example, a 4123 pattern might correlate with the misplacement of Releases, since it may be indicative of a fronted Release. The frequencies of these "high proportion" full strip patterns were correlated with the rates of several misplaced narrative categories (in Table 2). A positive correlation was found between pattern 1243 from all strips with misplaced Initiating Peaks, Peaks, and Releases (all *r*s < .610, all *p*s > .05). Pattern 2341 also showed a positive correlation with Establishers, $r(10) = .711$, $p < .05$. Out of all strip types, the 4123 pattern strongly correlated with all categories except Prolongations (all *r*s < .595, all *p*s < .05).

Within the EIPR strips, Peaks and Releases positively correlated with 1243 patterns, where they switched places (all *r*s > .580, all *p*s < .05). Positive correlations also appeared for switched Establishers and Initials (all *r*s > .580, all *p*s < .05) in a 2134 order. When the Establisher was moved to the end in a 2341 pattern, all narrative categories were positively correlated (all *r*s > .630, all *p*s < .05). Finally, when Releases were

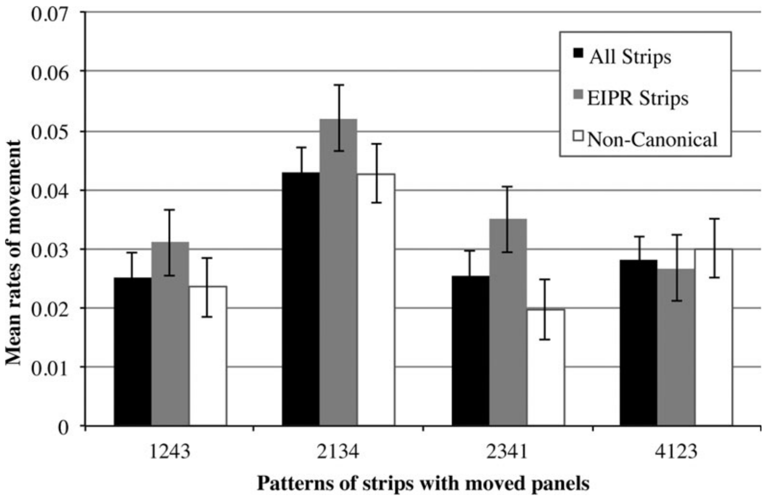

Fig. 5. The most frequent full strip patterns after movements take place.

fronted in a 4123 pattern, Establishers and Initials were positively correlated (all rs > .710, all ps < .05), while Peaks and Releases approached significance, $r(10) = .543$, $p = .068$ and $r(10) = .531$, $p = .076$, respectively. In Non-EIPR strips, only Peaks and Releases were correlated positively with the 1243 pattern, for both $r(10) = .582$, $p < .05$, while only Peaks showed a positive correlation with the 4123 pattern, $r(10) = .751$, $p < .01$.

2.3. Discussion

This experiment asked participants to reconstruct unordered panels into a coherent strip, from which we examined the rates at which categories were misplaced. Initials and Peaks seem the most stable, almost always being misplaced less than chance (except in EIPR sequences). Establishers, Prolongations, and Releases appear less constrained, and their rates of misplacement often hovered near or trended above chance (again, excluding EIPR sequences). These findings support a core versus periphery relationship between panels, consistent with discourse theories that distinguish between essential and non-essential parts of narratives (Halliday & Hasan, 1985; Trabasso & van den Broek, 1985).

In general, Peaks and Releases were misplaced less than categories that canonically appear toward the front of the sequence. Although Releases have some distributional flexibility, overall, categories at the end of a sequence seem to be more important than those at the front. Also of interest, Initiating Peaks were distinct in their misplacements. The rate of Initiating Peaks misplaced to the third position was the highest amount of any category into any position, yet movement into all other positions was marginal. Given that Position 3 is canonical for Peaks, it supports that these categories share similarities.

Similar distributions appeared between "peripheral" categories, reflected in the most frequent patterns of misordered strips. Incorrectly reconstructed Establishers or Releases were pushed to middle positions between Initials and Peaks. This occurred in the 1243 (EIRP) pattern, where a Release preceded a Peak, or in the 2134 (IEPR) pattern where an Establisher followed an Initial. This movement hints that these panels were functioning as Prolongations (themselves frequently misplaced). Establishers were also misplaced to the ends of sequences (2341—IPRE), while Releases were fronted (4123—REIP). This suggests that prototypical Releases and Establishers can be mistaken for one another. Finally, Prolongations were displaced very frequently to the front, but rarely to the end of sequences.

Altogether, these results suggest that the semantic content of panels appearing as Establishers, Prolongations, and Releases allows them to serve various roles within the sequence—at front, middle, or end—surrounding the more stable, core categories (Initial, Peak). These distributions further suggest an interaction between panels' content and narrative context, since both are necessary to determine the function of the panel. For example, Establishers typically appear at the front of a sequence and Releases at the end. If Establishers are frequently misplaced after Peaks, and Releases often shift to the front, evidently the content of those panels allows them to serve either role. This shows that content may have flexibility to play various roles in different parts of a sequence.

Finally, the salience of a canonical narrative arc is supported by the higher proportion of accurate reconstruction in EIPR strips compared with Non-EIPR strips. This meant canonical strips showed a lower proportion of moved categories. Canonical strips aid reconstruction because they provide a recognized pattern of narrative distribution. Despite this, the rates that categories were misplaced remained fairly constant between EIPR and Non-EIPR strips. These results support that the canonical pattern provides an advantage in recognizing the structure of a narrative by virtue of a global context (e.g., Mandler & Johnson, 1977; Rumelhart, 1975; Stein & Glenn, 1979; Thorndyke, 1977). However, the distribution of categories does not *only* rely on this global pattern, since the rates of misplaced categories did not differ between narrative patterns of strips.

While these data provide some evidence of complementary distribution for certain categories, much of the statistical support for such a conclusion was not fully significant or only trending in significance. Furthermore, these results reflect the choices made by participants constructing well-formed sequences, and they do not wholly test participants' intuitions for panels that were changed position out of their own control. Experiment 2 thus presented participants with sequences that reversed the positions of the original strip's panels.

3. Experiment 2: Panel reversals

In Experiment 1, certain narrative categories more readily changed positions in a sequence than others. "Core" categories, like Peaks, were less likely to move around than "peripheral" categories, like Establishers or Prolongations. Furthermore, these misplacements hinted at distributional similarities between Establishers and Releases—the same panels may be able to play both roles. However, these interpretations were based on participants' preferences for moving items around in a sequence, not judgments of manipulated sequences. Thus, Experiment 2 used sequences that reversed the order of certain pairs of panels. Participants controlled the pace of viewing each panel on a computer screen and we measured how long each panel stayed on the screen.

If the narrative role of a panel is solely determined by context (Jahn, 1997; Sternberg, 1982) or if narrative comprehension uses only linear semantic relationships (e.g., Magliano & Zacks, 2011; McCloud, 1993; Zwaan & Radvansky, 1998), viewing times should be equal for each category, regardless of their sequence position. For example, in this scenario, a Peak would be viewed at the same pace as an Establisher at Position 1, while an Initial at Position 4 would be viewed at the same rate as a Release.

In contrast, VNG predicts that viewing times for different categories would vary depending on their sequence position. A Peak moved to the first position (such as when reversed with an Establisher) of a sequence should be recognized as unusual, since completed actions prototypical of Peaks do not typically start sequences, nor do these semantic features usually map to Establishers (the category canonically in first position). While full "ungrammaticality" would need to be confirmed by the context of the subsequent panel (i.e., if a Release followed, it may be grammatical), having these semantic features of a culmination in the first position could lead to slower viewing times than other panels that more canonically fit, such as Establishers. Similarly, an Initial moved to the end (such as when reversed with a Release) should be equally poor, and thus viewed slower than other panels. In addition, if certain categories fall in complementary distribution, such as Establishers and Releases, we would expect that their reversal (Establishers to Position 4 and Releases to Position 1) would not change their viewing times. Such a result would support that the same panels can play different narrative roles in a sequence.

3.1. Methods

3.1.1. Stimuli

The same sequences from Experiment 1 were used in Experiment 2. However, all 180 strips were manipulated by reversing the position of two "critical panels." With each possible changing of positions (1-2, 1-3, 1-4, 2-3, 2-4, 3-4) as well as the original order, this created seven experimental sequence types, illustrated in Fig. 6. All sequence types were randomly distributed into seven counterbalanced lists so that each participant would only see each strip once. Within each list, strips were randomly ordered.

3.1.2. Participants

Twenty-one undergraduates from Tufts University (mean age: 19.81, 10 male, 11 female) participated in this study for compensation after providing their informed consent. All participants were self-defined "comic readers" and had an "average" comic reading expertise score (M = 15.97, SD = 6.8).

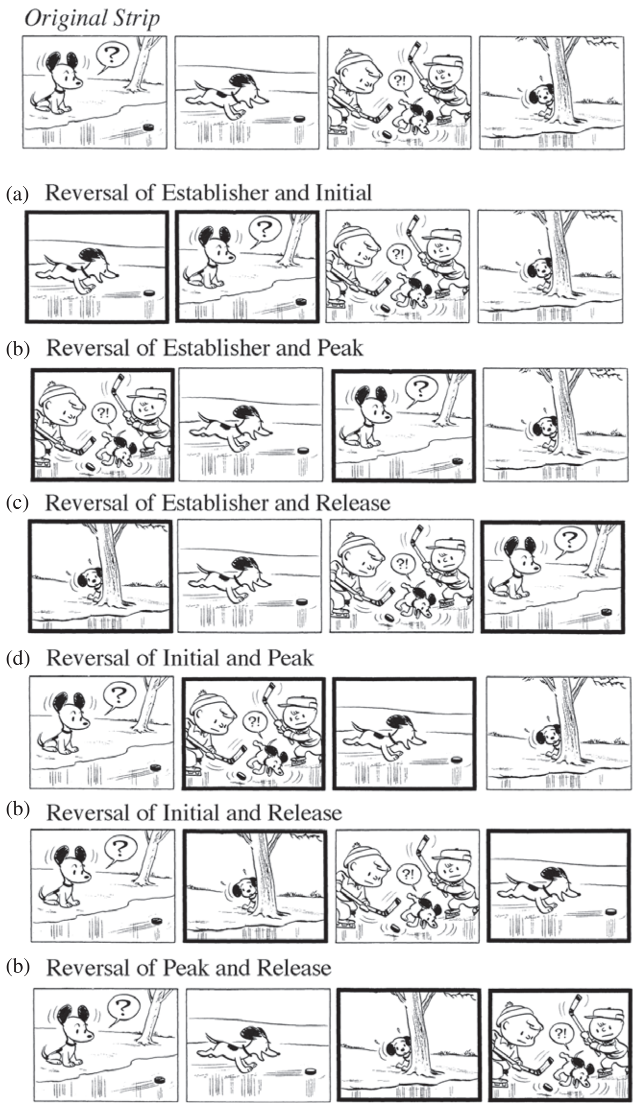

Fig. 6. Narrative sequences where two panels have been reversed in their positioning.

3.1.3. Procedure

Participants viewed each sequence on a computer screen and progressed through each panel of a strip one at a time by pressing a button under their own control. Trials began with a word "READY" at which point participants pressed a button to begin reading the strip. Four outlines of panels appeared on the screen, but content appeared in only one panel at a time. Panels progressed from the first (leftmost) to the last (rightmost) with each button press. After each button press, a 300 ms interstimulus interval delayed the onset of each subsequent panel to ensure that participants could not skip past panels faster than is possible to comprehend them (for example, by overexcited button pressing or accidentally holding down the button). After the last panel, a question mark appeared on the screen, where participants rated the coherence of the sequence using a 1 to 7 scale (1 = *hard to understand*, 7 = *easy to understand*).

3.1.4. Data analysis

Critical panels were identified as those that reversed position in a strip. Viewing times were measured for each button press assessing how long each panel remained on the screen. Only critical panels from canonical strips were analyzed, which ensured that all the same panels appeared at each position throughout a sequence, due to the canonical relationship between positions and categories. Non-canonical strips were treated as fillers.

Viewing times for critical panels at each sequence position (e.g., all categories at Position 1) were compared using four-way ANOVAS, along with four-way ANOVAS to compare the difference between viewing times of each category at each position (e.g., Establishers at Positions 1–4). An additional ANOVA examined participants' ratings of coherence for the whole strip. Follow up *t* tests further examined pairwise contrasts.

3.2. Results

3.2.1. Ratings

Ratings of coherence (see Fig. 7) differed significantly between all strips, $F(6, 120) = 8.75$, $p < .001$. The original canonical EIPR order was rated significantly higher than all manipulated orders (all ts > 3.15, all ps < .006). Intermediate ratings were found to E-I, E-P, I-P, and P-R reversals, which were significantly or near significantly higher than E-R and I-R reversals (all ts > 1.7, all ps < .088), but not from each other (all ts < .850, all ps > .405). E-R and I-R reversals were also not significantly different from each other, $t(20) = .088$, $p > .931$.

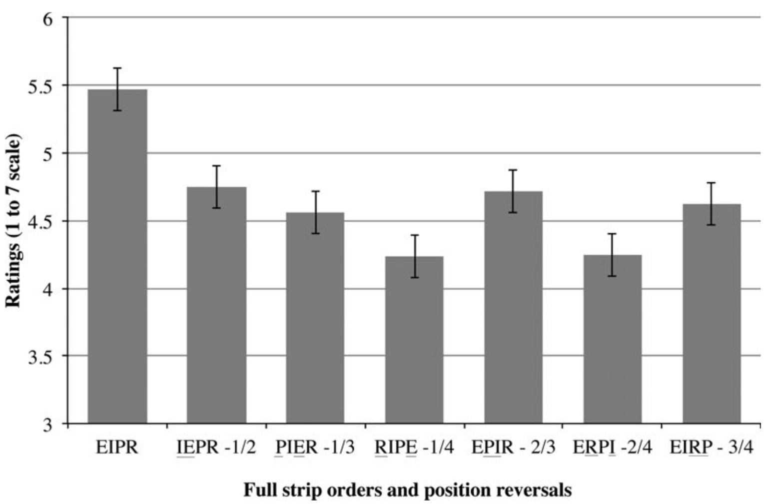

Fig. 7. Ratings of coherence to strips where panels had been reversed in their positions (underlined).

3.2.2. Viewing times

Viewing times for a position varied for different narrative categories (see Fig. 8). Significant differences between categories appeared at Position 1 and 4 (all Fs > 3.39, all ps < .05), but not at Positions 2 or 3 (all Fs < .666, all ps > .509). At Position 1, Peaks were significantly or near significantly slower than all other categories (all ts > 1.9, all ps < .068). However, Establishers, Initials, and Releases were not viewed at significantly different times. At Position 4, Peaks were viewed significantly slower than Establishers and Releases (all ts > 2.1, all ps < .05), but not Initials. Initials in turn were significantly slower than Releases, $t(20) = 2.18$, $p < .05$, but not Establishers. Establishers and Releases were not viewed at differing times.

In addition, all narrative categories differed in their viewing times across sequence position (all Fs > 7.4, all ps < .005). In all cases, viewing times were slower at Positions 1 and 4 than Positions 2 and 3.

3.3. Discussion

This experiment examined the distributional tendencies of narrative categories by measuring viewing times of panels that had been reversed in a visual narrative sequence. First, the canonical narrative pattern was comprehended faster and rated as more coherent than all manipulations from its structure at each panel position. However, at Positions 1 and 4, variance between the viewing times of categories supported the predictions of VNG.

Overall, panels were viewed the slowest at the beginning and ends of strips, regardless of narrative categories. Slower viewing times at the start of the sequence are consistent with previous research showing that people are slowest to read sentences at the outset of a verbal story (Haberlandt, 1980; Haberlandt, Berian, & Sandson, 1980; Mandler &

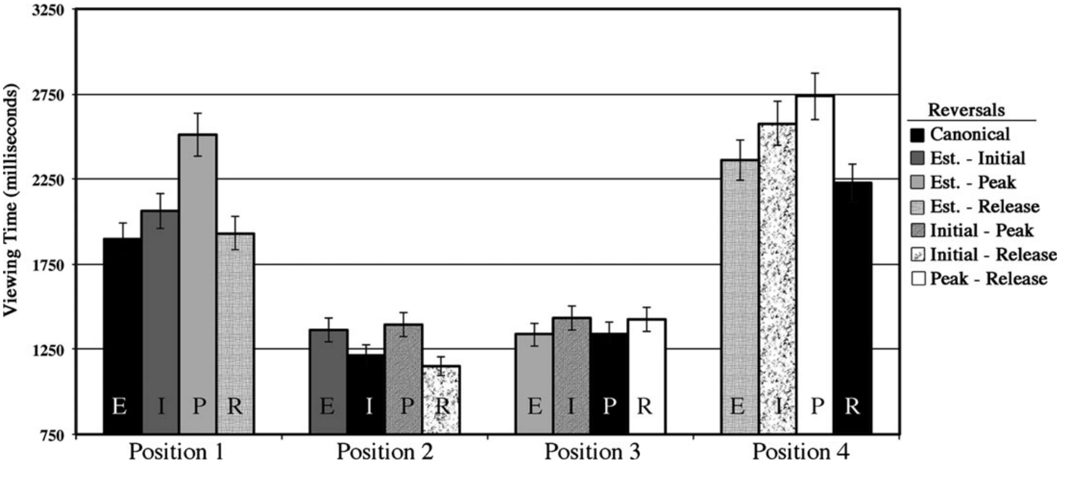

Fig. 8. Viewing times for panels that had been reversed in their position in a narrative sequence.

Goodman, 1982) or to view images at the outset of a visual narrative (Cohn & Paczynski, 2013; Cohn et al., 2012; Gernsbacher, 1983). Gernsbacher (1990) has argued that such slowing is part of a reader's process of "laying a foundation" of knowledge to be used in the subsequent discourse (which is similar to the function of an Establisher in VNG).

Among categories at Position 1, Peaks that moved to the front of a sequence (reversing with Establishers) were viewed the slowest. One explanation for this may be that, compared to the type of information in other categories, Peaks involve more active information about events, whereas other panels like Establishers do not. Thus, the content of Peaks requires more time to allow readers to integrate this event structure. However, content alone cannot explain this slowing because Peaks were not significantly slower than all other categories at every position. Thus, Peaks may also be viewed slower because of their context in the sequence. Starting a sequence with a culmination could be unusual and surprising, thereby resulting in longer viewing times (even if found to be "grammatical" by what follows it). Comparatively, starting a sequence with an Establisher or Initial should make more sense, and indeed these panels were viewed at comparable times, both faster than fronted Peaks. Interestingly, Releases were viewed at comparable times to Establishers and Initials, a point that will be discussed further below.

While viewing times at Positions 2 and 3 grew much faster, times again slowed down at Position 4. Previous literature has not consistently shown slowing at the end of verbal or visual narratives. However, given that participants knew these sequences were only four panels in length (seeing all four empty panels onscreen), this slowing down could reflect the anticipated "wrapping up" of reading the strip.

At this final position, Peaks and Initials were both viewed slower than other categories. In this case, the Peak would have followed its Release, while the Initial would have followed its Peak (which would be preceded by a Release). In other words, these strips would have presented information backwards. In both cases, this reversal would force a reader to reconsider the meaning of these panels given their order. These results support that Peaks, and to some extent Initials, remain relatively inflexible in their positioning in a sequence, similar to the findings of Experiment 1.

As expected at Position 4, both Peaks and Initials were viewed slower than Releases, which were in their canonical position. However, like at Position 1, Establishers were not viewed at different rates than Releases, despite also being presented "backwards" in the order of the sequence. These comparable times for Establishers and Releases at Positions 1 and 4 provide support that these panels play similar distributional roles in a sequence. At both positions, these panels also were viewed the fastest of all categories, again possibly reflecting the relative paucity of event structures represented in such panels—with fewer actions being shown, these panels require less time to integrate than other panels with more robust representations of events. This would suggest that the prototypically "passive" events in these panels could play acceptable roles at both the beginning and ends of sequences.

Nevertheless, strips where Establishers and Releases changed places were rated among the least coherent of all sequence orders—the other low-rated sequence having reversed Initials and Releases (*ERPI*). Thus, the two lowest rated sequences both featured Releases moved to the front two positions of the sequence. This suggests that, while panel-by-panel comprehension may remain similar for Establisher and Release panels, pushing "resolution" information forward in a sequence may change the meaning of the sequence in infelicitous ways. Such information may only be accessible once the entire strip has been read and integrated across panels, which is where the coherence rating would be given.

Overall, these results provide further evidence for some distributional tendencies for categories. However, they do not specifically inform us about categories' relative importance to the sequence. If peripheral panels can also be deleted with little effect on meaning, it would lend further support that they play a "non-essential" role for the sequence. In Experiment 3, a second Reconstruction task further examined the importance of certain narrative categories. However, an additional task measured participants' preferences for the expendability of certain panels by asking them to delete one panel from the sequence.

4. Experiment 3: Reconstruction-deletion

Previous research on verbal narrative has suggested that some parts of a narrative sequence are more expendable than others. Mandler and Johnson (1977) showed that the beginning of stories is acceptable to delete, so long as the context of the situation has been established in some other way. "Reactions" at the end of stories can also be acceptably deleted, but only when the beginnings of the story are left intact (Mandler, 1984); deleting other story endings creates distortions in the coherence and recall of the story. Also, the deletion of initiating actions may lead to a discourse with a more "surprising" narrative (Brewer & Lichtenstein, 1981). Thus, while some deletions are acceptable, they may rely on their relationship with other parts of the narrative.

This same interaction between intrinsic content and global context is predicted by Visual Narrative Grammar. Take for example Fig. 9, where categories have been deleted from Fig. 2a. In 9a, the deletion of the Peak leads to the strip no longer making sense. Why is Snoopy suddenly scared? Inference alone cannot fill in this information, since the Peak was an unexpected interruption. Similarly, deleting the Initial in 9b makes information feel noticeably missing; it jumps from an established relationship with a culminating event: how did Snoopy get there? On the other hand, deleting the Release in 9c renders the events of the sequence fairly complete, with less indication that something is missing. Nevertheless, the sequence ends abruptly and leaves the reader expecting that something should come next. Finally, the deletion of the Establisher in 9d has almost no impact on the sequences, just like deleting "establishing shots" from film sequences is hardly noticeable (Kraft, Cantor, & Gottdiener, 1991). Establishers set up characters and interactions, so their information is redundant with regard to later panels and their omission makes little difference to the sequence's meaning.

In Experiment 3, we again gave participants four unordered panels, and they arranged three of them while choosing one to exclude. We first looked at the rates that each

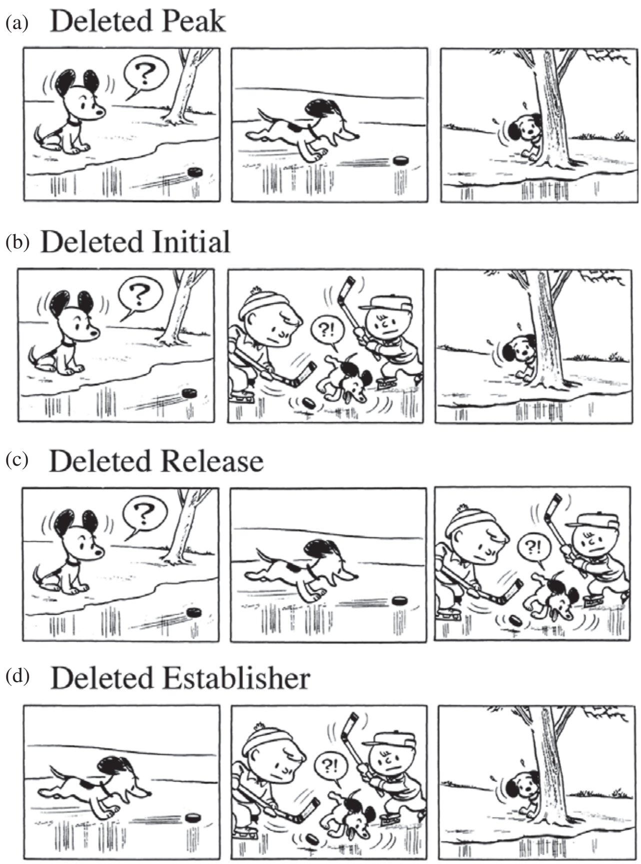

Fig. 9. Narrative categories deleted from a sequence.

original strip's positions were deleted, followed by which categories were deleted. If participants only use linear semantic relationships, inferences make the connections between panels (McCloud, 1993; Saraceni, 2000, 2003). Thus, panels in positions 2 and 3 should be less likely to be deleted, since they engage in two linear relations (1-2 and 2-3, 2-3 and 3-4), which would force more inference. Meanwhile, positions 1 and 4 should be more likely to be deleted, since they only have one linear relationship (1-2, 3-4). However, rates of deletion at positions 1 and 4 or positions 2 and 3 should be equal and not correlated with particular categories, since deletion would merely create further inference based on the number of linear relations between panels. In contrast, if only a global canonical pattern is used, panels should be deleted if they are expendable to the primary meaning of the canonical arc: Establishers and Releases (Mandler, 1984; Mandler & Johnson, 1977). Evidence against only using a top-down canonical schema would require categories in canonical and non-canonical strips to have the same patterns of deletion. This would support that semantic information about each panel influences their categorical role as well as the global schema.

Visual Narrative Grammar predicts such an interaction between the global structure and the roles of individual panels. Specifically, as described, if Initials and Peaks are more central to the comprehension of the narrative arc, participants should rarely delete them. In contrast, Establishers, Prolongations, and Releases should be deleted more often, since they are more peripheral to the narrative. Such results would support that categories play varying roles of importance to a narrative sequence and would be consistent across all types of strips.

4.1. Methods

4.1.1. Stimuli

The same 180 strips used in Experiment 1 were used in Experiment 3.

4.1.2. Participants

The same participants from Experiment 1 participated in Experiment 3.

4.1.3. Procedure

As in Experiment 1, participants were presented with four panels in a randomized cross formation in an open PowerPoint file. However, Experiment 3 added a black box in the upper right corner of the screen. Participants were asked to reconstruct the panels in an order that "made sense" using *only three panels*, and to place the unused panel into the black rectangle.

4.1.4. Data analysis

Reconstruction results were analyzed using the same methods as in Experiment 1, although accounting for the gap of the deleted panel. The position and category of the deleted panel were also recorded. Thus, if a panel from Position 2 was deleted, while the other panels retained their accurate relative placement it was recorded as "1-X-3-4."

Similar statistical measures were used in Experiment 3 as in Experiment 1. For the Reconstruction task, misplaced narrative categories were individually compared against chance using t tests. However, because one panel was deleted, chance was recalculated by subtracting the number of deletions for a particular category from the total possible for each subject. Then, across participants, the sum of these non-deleted panel frequencies was divided by the sum of all non-deleted panels. For the Deletion task, we compared the percentages of deleted positions and narrative categories using a chi-squared analysis, followed by pairwise analyses using t tests. Deleted panels were also compared against chance using t tests. The probability for deletion of each position was found by averaging each participant's mean rate of deletion.

Again, one-way ANOVAS used "task order" as a between-subjects factor to examine whether the order in which the Deletion Task appeared relative to the tasks from Experiments 1 and 4 influenced the results. These ANOVAS were carried out for each individual position or category under analysis.

4.2. Results

4.2.1. Reconstruction

The Reconstruction task in Experiment 3 yielded an even higher proportion of correct sequences than Experiment 1; 86% of whole strips were accurately reconstructed, while 90% of individual panels maintained correct relational positions. As in Experiment 1, narrative categories were misplaced in non-canonical positions (M = .051, SD = .04) significantly more than canonical positions (M = .012, SD = .014), $t(11) = -4.28$, $p < .005$.

No main effects appeared between the rates that different narrative categories were misplaced (excluding Prolongations, which had a rate of misplacement of 0). Also, no main effects appeared when analyzing the influence of task order (all Fs < 1.66, all ps > .120). On the whole, categories were misplaced less frequently than in Experiment 1, with greater variability: Releases were misplaced the most (M = .13, SD = .13), followed by Establishers and Initials (M = .11, SD = .09), and then Peaks (M = .09, SD = .09) and Initiating Peaks (M = .08, SD = .14). Establishers, Initials, and Peaks were all misplaced less often than chance (all ts > 4.4, all ps < .005), while only Releases were misplaced more often than chance, $t(11) = -3.20$, $p < .01$.

4.2.2. Deletion task: Positions

Overall, participants chose to exclude certain positions over others, $\chi^2(3,357) = 60.69$, $p < .001$. Positions 1 (40%) and 4 (28%) were deleted far more often than Position 2 (19%) and Position 3 (13%). No significant difference in task order appeared for the rates of deletion for each position (all Fs < 2.5, all ps > .136). The rates of these positions were also compared against the 25% chance that they would be excluded. Position 1 was deleted higher than chance (.25), $t(11) = 3.53$, $p < .01$ and Position 3 lower than chance, $t(11) = -9.55$, $p < .001$. Positions 2 and 4 did not differ significantly from chance (all ps > .130).

4.2.3. Deletion task: Categories

Categories were also omitted in differing proportions, $\chi^2(5,360) = 153.87$, $p < .001$. Establishers (31%), Initials (24%), and Releases (26%) were deleted far more often than Peaks (12%), Initiating Peaks (5%), or Prolongations (3%). Task order had no significant impact on rates of deletion for categories (all Fs < 1.94, all ps > .199). Because categories were not evenly distributed throughout the strips of the experiment, preferences for deletion become clearer when looking at the percentage of each category that was deleted, summarized in Fig. 11. Establishers, Prolongations, and Releases were deleted above chance, while Initials, Initiating Peaks, and Peaks fell below chance. However, these rates of deletion were only significant for Establishers, $t(11) = -3.57$, $p < .01$, and Peaks, $t(11) = 7.37$, $p < .001$.

Next, we asked whether deleted categories correlated with particular positions. Deleted Establishers positively correlated with Position 1, $r(10) = .894$, $p < .001$, Initials with Position 2, $r(10) = .894$, $p < .001$, and Initiating Peaks with Position 1, $r(10) = .707$, $p < .001$. Peaks were correlated negatively with Position 2, $r(10) = .686$, $p < .05$, but positively with Position 4, $r(10) = .696$, $p < .05$. Releases were also correlated negatively with Position 2, $r(10) = -.705$, $p < .05$, and positively with Position 4, $r(10) = .945$, $p < .001$.

Further analysis looked at the deleted categories from EIPR and Non-EIPR strips. Categories were deleted at different rates across all EIPR strips, $\chi^2(3,177) = 50.14$, $p < .001$, summarized in Fig. 10. Establishers and Releases were chosen for deletion far more often than Initials and Peaks. Compared to the 25% chance that they would be deleted, only Releases did not have mean rates of deletion significantly different than chance. We then looked at the percentage of each category that was deleted. The same general patterns maintained: more Establishers and Releases were deleted, and less Initials and Peaks were deleted, summarized in Fig. 11. Of these, Establishers were deleted significantly more than chance, $t(11) = 3.22$, $p < .01$, while Initials, $t(11) = -2.97$, $p < .05$, and Peaks, $t(11) = -5.80$, $p < .001$, were below chance.

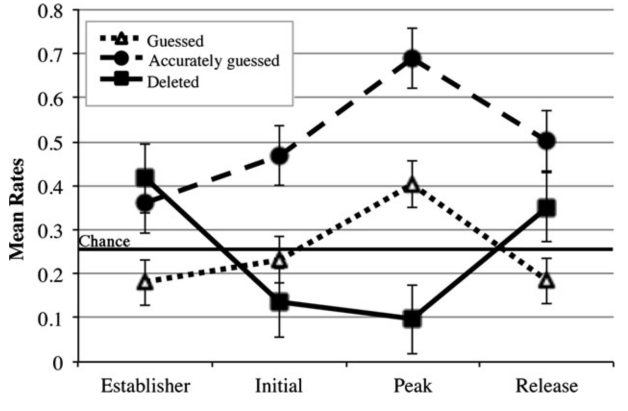

Fig. 10. Mean rates for EIPR strips of panels chosen to be deleted for Experiment 2 and rates of Guesses and Accurate Guesses for deleted panels from Experiment 3.

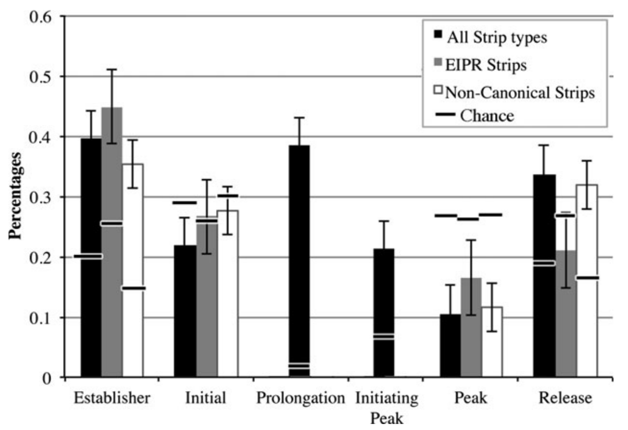

Fig. 11. Percentages of each category chosen to be deleted.

Categories were again deleted at different rates in Non-EIPR strips, $\chi^2(5,183) = 58.41$, $p < .001$. Of all panels deleted in Non-EIPR strips, Establishers (21%) and Initials (34%) were deleted the most often, followed by Releases (18%) and Peaks (12%). The least often deleted categories were Prolongations (6%) and Initiating Peaks (9%). All categories were deleted significantly less often than chance (all *t*s > 2.6, all *p*s < .05), except for Prolongations which were still below chance, just not significant, $t(11) = 1.620$, $p = .134$.

When analyzing the percentage of panels within each category chosen to be deleted in Non-EIPR strips, the same general trends persisted as across all strip types (summarized in Fig. 12): Prolongations were deleted the most overall, followed by Establishers, Releases, and Initials. In contrast, Initiating Peaks had the second lowest percentage of deletion, behind only Peaks. All categories were deleted at rates significantly higher than chance (all *t*s < −2.22, all *p*s < .05) except Peaks, which were deleted at lower than chance, $t(11) = 3.68$, $p < .005$, and Initials, which were nearly equal with chance, $t(11) = -.024$, $p = .981$.

4.3. Discussion

This experiment again examined how unarranged panels were reconstructed into a linear order, but also probed which panels participants choose to delete. The Reconstruction results in Experiment 3 repeated the findings from Experiment 1, although mean rates of misplacement were comparatively lower. Releases, Establishers, and Initials were misplaced only slightly more frequently than Peaks. However, these differences in movement were marginal and only Releases were moved more than chance. This reduced rate of misplacement likely reflects the increased probability of accurate reconstruction due to

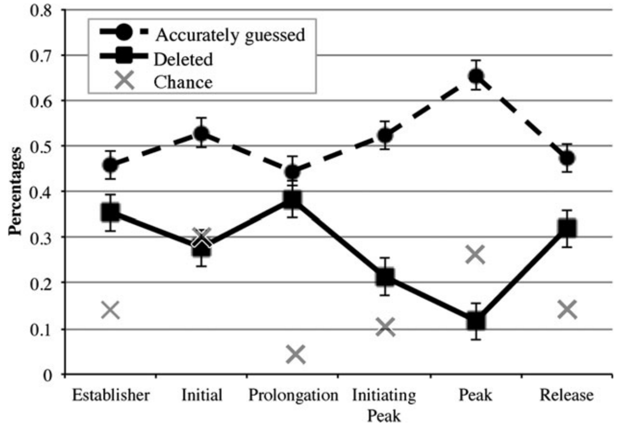

Fig. 12. Percentages of categories in non-EIPR strips of panels chosen to be deleted for Experiment 2, and rates of Accurate Guesses for deleted panels from Experiment 3.

the Deletion task: Since one panel was eliminated, only three panels—not four—needed to be ordered.

The Deletion task reflected the same pattern of essential and non-essential categories as the Reconstruction task of Experiment 1. Overall, panels were deleted from positions on the outside of the sequence (Positions 1 and 4) more often than they were from the inside of the sequence (Positions 2 and 3), though only Position 1 was significantly higher than chance, while Position 3 was significantly lower than chance. These relative rates of deletion for positions aligned with the canonical positions for narrative categories. Out of all deleted categories, Establishers, Initials, and Releases were deleted the most often, with Peaks among the least. Also, the percentage of deletions of Establishers, Releases, and Prolongations was higher, while the percentage of deleted Peaks, Initials, and Initiating Peaks was lower. These patterns are consistent with previous research showing that beginnings of verbal and film narratives are more acceptable to be omitted (Kraft et al., 1991; Mandler, 1984; Mandler & Johnson, 1977) as well as some endings (Mandler, 1984; Mandler & Johnson, 1977). Nevertheless, only Establishers were deleted more than chance and Peaks less than chance. These results suggest that the most expendable category shows the least information about events (Establishers), while the least expendable carries the most information about events (Peaks).

These trends persisted in EIPR and Non-EIPR sequences. Establishers again were deleted at significantly higher rates than chance in EIPR and Non-EIPR sequences, further supporting their status as less essential to the sequence. The percentage of deleted Releases and Prolongations was also above chance in Non-EIPR sequences, suggesting that they too are less essential. In contrast, Initials and Peaks were omitted significantly less than chance in EIPR sequences, and Peaks less than chance in Non-EIPR sequences. Thus, Peaks and Initials appear more essential to a sequence. Additionally, Initiating Peaks were deleted at comparable rates to, or directly between those of, Initials and Peaks. This supports that these panels share features of Initial and Peak types. Overall, these findings support that Peaks and Initials create a core of the narrative that is surrounded by other peripheral categories.

If the "peripheral" categories of a sequence do carry less relative importance to the sequence, their deletion should be less noticeable than the deletion of the "core" categories, which would highly impact a sequence's coherence. Experiment 4 explored participants' intuitions of coherence for sequences of categories, using the reverse logic of the Deletion task.

5. Experiment 4: Deletion recognition and coherence rating

Just as choosing to delete panels can inform us about their importance to a sequence, so can recognizing the location of a deleted panel. In many cases, the ellipsis of information from a narrative demands that a reader infer what is missing. For example, it should be apparent that the sequence in 9a is missing key information (a Peak), while the missing Establisher in 9d has almost no impact. Thus, Peaks may demand inference if they are deleted, but Establishers may not. Theories of sequential images have focused on inference as the driving force behind comprehension of *all* adjacent images, not just particular omissions, as in McCloud's (1993) theory of "panel transitions." Saraceni (2000, 2003) operationalized this inference as driven by the degree to which adjacent panels repeat characters and relate to a common overarching semantic field. Thus, the predominant view of sequential image comprehension places inference as driving the comprehension of all adjacent panels.

Some experimental evidence has looked at the recognition of deleted information in sequential images. Nakazawa and Nakazawa (1993) asked participants of various ages to identify the contents of omitted panels in comic strips. Across age groups, no children in kindergarten or first grade could accurately determine what was missing, although older children progressively became better at recognizing the missing information. College-aged adults, who read comics more often, more accurately inferred the missing contents than younger children and older adults (Nakazawa, 2004). These results suggest inference for missing content increases with age and experience reading comics.

Inference in verbal discourse also has a rich literature (for review, see van den Broek, 1994; Zwaan & Rapp, 2006). One line of research has focused on the causal connections found between sentences. Faster reading times and more accurate memory have been found to sentences with stronger causal connections to prior sentences than those with weaker connections (Keenan, Baillet, & Brown, 1984; Myers, Shinjo, & Duffy, 1987), implying that the generation of inference depends on the strength of the causal relationship. Creating such inferences requires the interactions of causal properties of events and actions in the absence of that information (Singer, Halldorson, Lear, & Andrusiak, 1992). Relative to VNG, such connections would support that the omission of Initials and Peaks would demand more inference, since these panels contain most of the causal information about events (i.e., in prototypical cases preparatory actions for Initials and completed actions for Peaks).

Story grammar research has suggested that knowledge of missing information is motivated not just through causal relationships, but through reference to a global canonical narrative structure, a position strongly supported by film theorists (Bordwell, 2007; Bordwell & Thompson, 1997). In experiments where participants were asked to recall information moved away from the canonical narrative order, readers falsely remembered it in the canonical order, not the surface positions where elements actually appeared (Mandler, 1984; Nezworski, Stein, & Trabasso, 1982). These findings indicated that readers expect particular information to be located in the narrative based on a canonical pattern, not just on the causal connections or inferences between discourse units.

Given this background, Experiment 3 directly complemented the Deletion task by presenting participants with a three-panel strip and asking them to identify where a panel had been deleted, and then rate the coherence of the sequence. We first asked how accurate on the whole participants would be at guessing the location of missing panels. If participants only used a global context to inform their guesses, guess for canonical sequences should be more accurate than for non-canonical sequences. This would arise because the canonical structure would guide the knowledge of what would be missing from the sequence, while non-canonical sequences would lack such cues. However, if narrative categories play roles beyond their sequence context and are motivated by their content, similar accuracy of recognition should occur between canonical and non-canonical sequences.

Next, if participants use only linear panel relationships, we would predict that all positions would be guessed in equal proportion. Here, the inference demanded by a missing panel should be no greater than the inference demanded by any other juxtaposition, and thus all positions should be equally likely to be missing. On the other hand, if a global pattern governs comprehension, it predicts that—if participants assume crucial information is missing—Positions 2 and 3 would be guessed more than any other, since they are prototypical of Initials and Peaks, which carry the most important information for the sequence.

We next asked how accurate participants were at guessing the original location of missing panels based on their positions and categories. Again, because the panel transition approach assumes that inference bridges the gaps between *all* panels, no appreciative difference should arise between the rates of recognition between different types of panels. Any missing panel should have the same inferential demand as any other, regardless of narrative category. Alternatively, if only the strength of causal relationships dictates the recognition of missing panels (e.g., van den Broek, 1994; Keenan et al., 1984; Myers et al., 1987), panels with content featuring greater causal connections would be recognized more (i.e., Initials, Peaks, Releases) than those with less causal connections (i.e., Establishers, Prolongations). However, such a view would make no predictions about differences between the global relationships of canonical and non-canonical sequences. With reliance solely on a global canonical narrative, the recognition rates of individual categories would change based on the pattern of the strip. In particular, missing categories in canonical strips would be more accurately recognized than those in non-canonical strips.

VNG predicts a combination of both of these latter perspectives. The properties of panels' contents (such as causal relations) should influence a panel's rate of recognition. However, schematic knowledge of a global canonical narrative structure should also facilitate the recognition of missing components. Participants' ability to detect missing categories of a narrative should complement their choices for deletions in Experiment 3. For example, missing Establishers would be recognized less often for the same reason as they were deleted more. Peaks would be the opposite: their deletions should be recognized more often for the same reasons as they were deleted less. Finally, strips missing core information (Initials, Peaks) would be rated as less felicitous than strips missing peripheral information (Establishers, Prolongations, and Releases).

Finally, if participants only use linear relations between panels, all sequences should be rated as highly coherent, since all panel relations require inference, and missing panels would just warrant the same type of inference demanded by any other juxtaposition. Thus, the coherence of whole strips would not differ based on particular omissions. In contrast, reliance solely on a global pattern would assume that canonical sequences with missing panels that are less essential (i.e., Establishers, Releases) would be rated as higher than in non-canonical sequences. This would arise because the sequence would get its meaning from the global context. However, if sequences with these same missing categories receive similar ratings of coherence in non-canonical sequences, participants may be using cues from the panels themselves, beyond the global context, as predicted by Visual Narrative Grammar.

5.1. Methods

5.1.1. Stimuli

The same 180 strips used in Experiments 1 and 3 were used in Experiment 4.

5.1.2. Participants

The same participants from Experiments 1 and 3 participated in Experiment 4.

5.1.3. Procedure

In an open PowerPoint file, participants were presented with a three-panel strip, which they were told was originally a four-panel strip where one panel was deleted. Participants were asked to drag a red square into their assumed location of the missing panel. They were then asked to rate the strip for "how easy it is to understand" on a 7-point scale (1 = *very hard to understand*, 7 = *very easy to understand*).

5.1.4. Data analysis

We analyzed participants' ratings of coherence, the rates of guesses for panel positions, and rates of accurate guesses for categories and positions. A chi-squared analysis compared rates of accurately recognized categories and positions as well as guessed positions. These were followed by *t* tests of pairwise relationships. *t* tests compared guesses and accurate recognition against chance for positions (.25) and narrative categories (see Table 1). ANOVAS were used to compare coherence ratings of narrative categories and positions. Again, a Pearson correlation was used to compare rates of recognized panels and guesses with ratings.

Finally, one-way ANOVAS using "task order" as a between-subjects factor examined whether the order in which the Recognition Task appeared relative to other experimental tasks from Experiments 1 and 3 influenced the results. These analyses were initially carried out for the combined results (EIPR and Non-EIPR), with follow up analyses performed for significant findings at individual strip patterns. Significant findings were followed by *t* tests for individual relationships.

5.2. Results

5.2.1. Recognition task

Overall, the rate of accurately recognizing the location of a missing panel was only 51%. A similar rate was maintained for EIPR strips (50%), and Non-EIPR strips (53%), a difference that was not significant.

5.2.1.1. Positions: Whether accurate or not, rates of guesses for the positions of deleted panels differed greatly, $\chi^2(3,1422) = 130.6$, $p < .001$. Participants clearly guessed Position 3 more than any other position (38%), followed by Position 2 (22%), Position 4 (20%), and Position 1 (19%). Position 3 was guessed significantly higher than chance, Position 1 and Position 4 significantly below chance, and Position 2 trending below chance (all *t*s > 1.85 or <−5.17, all *p*s < .09).

The rates of guesses for positions of EIPR strips are reported below, with discussion of narrative categories. Non-EIPR strips had significant differences between rates of guesses, $\chi^2(3,725) = 45.96$, $p < .001$. Position 3 was still guessed the most (36%), followed by all other positions (Position 4 = 22%, and Positions 1 and 2 = 21% each). Position 2 was guessed below chance, Position 1 trended below chance, and Position 3 above chance (all *t*s > 4.17 or <−1.95, *p*s < .077). Position 4 was not significant, $t(11) = -1.45$, $p = .175$.

Rates of accurate guesses for positions differed as well, $\chi^2(3,1422) = 235$, $p < .001$. Participants were most accurate at recognizing Position 3 (67%), followed by Position 4 (53%), Position 2 (46%), and Position 1 (41%). All positions were guessed significantly above chance (all *t*s > 3.5, all *p*s < .01). Accurate recognition across positions in Non-EIPR strips was significant, $\chi^2(3,385) = 12.85$, $p < .01$. Again, Position 3 was most accurate (66%), but followed by Position 4 (56%), Position 2 (45%), and Position 1 (44%). Accuracy for these guesses was significantly above chance for all positions (all *t*s > 4.5, all *p*s < .005). Position 3 was highest, followed by Position 4, 2, and then 1. Rates of accurate guesses for EIPR strips will be discussed below, with category information.

Significant differences arose between task orders for the rates of accurate guessing of Position 1, $F(2,11) = 7.56$, $p < .05$, but not any other position, accurate or not (all *F*s < 2.88, all *p*s > .107). Follow-up analysis confirmed that a lower rate of accuracy occurred for Position 1 when participants viewed Task order 2 ($M = .25$, $SD = .08$), in which the Recognition task was the first task performed, than for Task order 1 ($M = .44$, $SD = .11$) or Task order 3 ($M = .53$, $SD = .12$; all *t*s > 2.7, all *p*s < .05), where the Recognition task followed other tasks (and where rates did not differ from each other).

5.2.1.2. Categories: Rates of guesses for panels in EIPR strips—accurate or not—differed between all positions/categories, $\chi^2(3,697) = 90.97$, $p < .001$, summarized in Fig. 10. Significant differences were found below chance (.25) for Position 1/E and Position 4/R, and above chance for Position 3/P (all *t*s > 4.3 or <−2.6, all *p*s < .05). Position 2/I was not significantly below chance, $t(11) = -.869$, $p = .409$.

Accurately guessed categories differed between all missing panels, $\chi^2(5,738) = 330.83$, $p < .001$. The rates of recognition for missing Peaks (17%) and Initials (14%) were twice as high as Releases (9%), Establishers (8%), Prolongations (1%), and Initiating Peaks (3%). All categories were guessed as missing significantly below chance (all *t*s > 5.0, all *p*s < .001). Looking at the rates of recognition within categories provided further insight. Two-thirds of all omitted Peaks were accurately recognized as missing (66%), roughly half of Initials (50%), Prolongations (44%), and Initiating Peaks (52%) were accurately guessed, while just four in 10 Establishers were recognized (40%).

Accurately guessed categories in EIPR strips also differed from each other, $\chi^2(3,353) = 19.09$, $p < .001$. These rates are summarized in Fig. 10. All categories were significantly above or trending above chance, set at .25 (all *t*s > 1.9, all *p*s < .075). Peaks were recognized far more often than any other category, followed by Releases and Initials, and finally Establishers. Recognized categories from Non-EIPR strips were also significantly different, $\chi^2(5,385) = 167.2$, $p < .001$. Initials (16%) and Peaks (17%) were accurately recognized more than double the rate of any other category—Establishers (6%), Prolongations (2%), Initiating Peak (5%), and Releases (5%). However, while Initials were also recognized above chance, only Peaks were significantly above chance, $t(11) = 3.364$, $p < .01$. Proportionally, 65% of Peaks were accurately noticed when missing, while all other categories hovered near 50% (summarized in Fig. 12).

Significant main effects were found for task orders in the accuracy of guessing missing Establishers, $F(5,55) = 8.06$, $p < .05$, but not for any other category (all *F*s < 3.5, all *p*s > .072). Again, this difference arose because Establishers were less accurately guessed in Task order 2 ($M = .04$, $SD = .02$), than Task order 1 ($M = .09$, $SD = .009$) or Task order 3 ($M = .10$, $SD = .03$; all *t*s > 3.2, all *p*s < .05). Task orders 1 and 3 did not differ. Follow ups revealed that Establishers in EIPR strips upheld this same pattern for Task order 2 being less accurate than other task orders (all *t*s > 3.3, all *p*s < .05), but again Task orders 1 and 3 did not differ.

5.2.2. *Ratings of coherence*

On a 7-point scale, participants rated strips to be more coherent when they also accurately recognized the location of the missing panel (mean rating: 5.73, $SD = .77$) than when their guesses were not accurate (mean rating: 5.68 $SD = .69$). Ratings were also higher for EIPR strips when participants were able to accurately recognize categories (accurately guessed: $M = 5.79$, $SD = .66$, not accurately guessed: $M = 5.65$, $SD = .68$), although the reverse was true in Non-EIPR strips (accurately guessed: $M = 5.73$, $SD = .73$, not accurately guessed: 5.64, $SD = .9$). However, none of these differences were found to be significant.

5.2.2.1. Positions: Significant main effects between ratings of positions, $F(3, 33) = 3.518$, $p < .05$, although no significant pairwise relations were found. Task order showed no significant effects (all *F*s < .421, all *p*s > 669). Ratings were highest for strips that deleted panels in Position 1 ($M = 5.9$, $SD = .6$), decreasing slightly for Position 2 ($M = 5.8$, $SD = .6$), and were lower for Position 3 ($M = 5.5$, $SD = .9$) and Position 4 ($M = 5.5$, $SD = 1.0$). Ratings for deleted positions in Non-EIPR strips trended between all positions, $F(3, 33) = 2.59$, $p = .069$. Mean ratings again showed a steady decrease in coherence across the sequence for missing panels: Position 1 ($M = 5.9$, $SD = .6$), Position 2 ($M = 5.77$, $SD = .7$), Position 3 ($M = 5.6$, $SD = .7$), and Position 4 ($M = 5.4$, $SD = 1.3$).

5.2.2.2. Categories: Ratings for deleted narrative categories across all strips did not yield significant main effects, $F(5, 55) = .768$, $p = .577$, and was not influenced by task order. Ratings are as summarized in Fig. 13, strips with missing Establishers, Prolongations, and

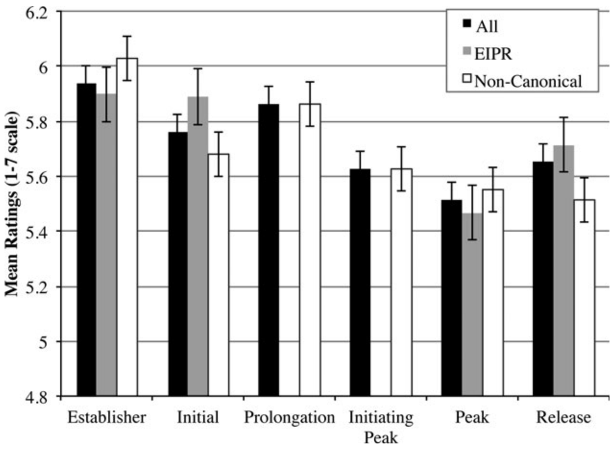

Fig. 13. Mean ratings for the felicity of strips with deleted narrative categories in various types of strips.

Initials were rated more coherent than strips with missing Peaks, Initiating Peaks, and Releases. Differences between ratings of EIPR strips with missing categories trended toward significance, $F(3, 33) = 2.68$, $p = .063$. Task order showed no significant effects (all Fs < .735, all ps > .506). Mean ratings for these strips maintained the same pattern as other strips, with ratings to missing Establishers and Initials remaining fairly high and Peaks fairly low. However, ratings for missing Releases fell between those for Establishers/Initials and Peaks. These data are depicted in Fig. 13. Non-EIPR strips showed no significant difference in the ratings for deleted narrative categories, $F(5,55) = 1.03$, $p = .408$. However, the general proportion of ratings was similar to those in all strips: higher ratings for strips with missing Establishers or Prolongations, low ratings for missing Peaks or Releases, and medial ratings for missing Initials or Initiating Peaks.

5.3. Discussion

This experiment examined how well participants could recognize narrative categories when they were deleted from a sequence, and their assessment of what impact those deletions had on the coherence of the sequence. With only a 50% rate of recognition of missing panels overall, participants appeared marginally proficient at recognizing the locations of deleted panels. However, the rates of recognition between deleted positions and categories differed greatly, indicating that recognition depended largely on what was deleted. More careful examination of the results suggests an interplay between a canonical narrative structure and individual categorical roles.

First, the rates of guesses for panel positions imply that people rely on expectations from a canonical narrative structure. For example, Position 3 was guessed more than any other position and with higher accuracy, possibly because it is the canonical position for a Peak. Conversely, Position 1—prototypical of Establishers—was guessed less and least accurate. Thus, when told something was missing, participants suspected it would be in the position of important (Peaks) rather than less important information (Establishers), even when the actual Peak was depicted. This supports that participants use the canonical pattern in their expectations about a visual narrative.

Despite this global structure, participants did recognize deleted categories in differing proportions. No matter the strip types, Peaks were most often recognized as missing, while Establishers and Prolongations were rarely recognized. Initials and Releases had moderate rates of recognition. These results suggest that Establishers contain information that is less salient for the sequence because they are not missed when gone. Conversely, Peaks make a large impact when missing because their ellipsis is noticed. These results are consistent with the expectations from research of causal inferences (van den Broek, 1994; Keenan et al., 1984; Myers et al., 1987), since panels with the most enriched content were inferred more than those with less distinct content.

The ratings of coherence further suggest that panels play different roles in relation to the overall narrative structure. Ratings were highest for panels that were missing from the beginning of sequences, but progressively grew worse as panels were missing later in sequences. Strips with missing Peaks were rated the lowest, no matter the sequence pattern. Strips with missing Releases were also rated very low, especially in non-canonical strip patterns, as were Initiating Peaks. In contrast, strips with missing Establishers, Initials, and Prolongations were all rated as fairly coherent. Nevertheless, ratings for accurately recognized panels in canonical strips were higher than those in Non-EIPR strips, showing a preference for the canonical pattern beyond just the causal features depicted in panels.

Ratings for strips with missing categories reflected the same preferences for recognizing deleted narrative categories. Strips with highly recognized missing panels (Peaks) were rated as less coherent, while strips with panels that were not recognized (Establishers, Prolongations) were rated as more coherent. Deleted Initials and Releases were recognized with moderate accuracy, but strips missing Initials were rated fairly high, while strips missing Releases were rated very low. These trends may reflect aspects of the prototypical content of these categories. Initials and Releases both depict information directly relevant to the events of a sequence, and thus are recognized when missing. However, because Initials lead up to the Peak, this content may be inferable from the Peak itself. In contrast, Releases prototypically depict the endpoint or aftermath of the narrative or events—which, because they are at the end, are not inferable from other panels. This finding is consistent with research on path actions which shows that sources and starting points are more salient than goals and endpoints of paths, whether it is in language (Lakusta & Landau, 2005) or perception and attention (Regier, 1996, 1997).

Finally, it is also worth addressing the significant results of “task order” with regard to guesses for Position 1 and Establishers being less accurate for participants who performed the Recognition task prior to the other tasks. By inference, this suggests that participants who preceded this task with Experiments 1 and 3 became more accurate at guessing these missing panels because they were able to learn the structure of the strips across tasks. However, accuracy for other positions and categories did not receive this benefit, and no effects of task order appeared for other experimental tasks. Also, no differences arose in the accuracy for Establishers when the Recognition task appeared as the second task versus the third task, suggesting no incremental process of learning across subsequent tasks. Furthermore, even if participants became specifically sensitized to Establishers' role in strips, the rates of recognition shown by participants who "learned" this structure were still lower than nearly all other categories. Thus, if learning did occur, the reported poor accuracy shown to recognizing Establishers reflects an inflated effect, and without learning, this accuracy for Establishers may have been even worse.

Taken together, these results further support that Initials and Peaks are core categories of a sequence, while Establishers, Prolongations, and Releases are less essential. However, these categories interact with a canonical structure that allows participants to make expectations about the properties of a narrative sequence.

6. General discussion

In four experiments, we examined the distributional trends of narrative categories theorized in Visual Narrative Grammar. Consistent distributional patterns arose for individual categories in complementary ways across experiments, indicating that panels have at least some functional identity motivated by their semantic content within a larger global context. Below, we further explore the characteristics and implications of these results.

First, these results suggest that a narrative structure has both core and peripheral components. This is consistent with similar distinctions made by theories that emphasize the contrast between essential and non-essential information in a verbal discourse (Black & Bower, 1980; Halliday & Hasan, 1985; Trabasso & van den Broek, 1985; Trabasso et al., 1984). These results imply that this contrast applies to narrative *categories* in an overall structure. Peaks and Initials appeared more central to the narrative and were misplaced less, deleted less, recognized when missing more, and viewed longer when out of position. In contrast, Establishers, Prolongations, and Releases appear more peripheral; they were moved more, deleted more, recognized when missing less often, and viewed at comparable times in reversed positions. That is to say, the core and peripheral aspects of narratives play particular roles in relation to a global context.

While these experiments support the roles of individual categories, they also are consistent with research showing that a canonical pattern aids in the comprehension of narrative sequences (e.g., Mandler & Johnson, 1977; Rumelhart, 1975; Stein & Glenn, 1979; Thorndyke, 1977). In Experiment 1, canonical strips were reconstructed at higher rates than non-canonical strips, while in Experiment 2 nearly all categories in canonical positions were viewed faster than those out of their original position. Experiment 4 showed that participants guessed missing panels at Position 3 more than any other panel—the canonical Peak position. Also, canonical strips with missing panels were given higher ratings of coherence than non-canonical strips. These results support that a canonical structure plays a role in the understanding of narratives.

Despite this evidence supporting a canonical pattern, it is not the only possible sequence. Narrative categories can appear in non-canonical positions as well, given the constraints of the global constituent structure. Overall, narrative categories did not differ substantially in their distributional traits (movement, deletion, and recognition) between canonical and non-canonical strips. This implies that, although an entrenched canonical structure may underlie comprehension, individual panels also motivate their behaviors in a sequence. These findings support that the content of panels and their role in a global context are both important in the structure of narratives.

6.1. Traits of narrative categories

Individual narrative categories showed distributional tendencies beyond the broad distinction between core and peripheral functions. As hypothesized by VNG, the differing distributions for these categories can aid in identifying and characterizing these categories. That is, if a panel's categorical nature is unknown, these tendencies provide operational diagnostics for recognizing its narrative category: What position in the sequence does it gravitate toward? Can it be deleted with little recourse on coherence? Will it be recognized if it is missing? Can it be moved to other positions and retain its coherence? Below, we describe these tendencies.

6.1.1. Peaks

Across all experiments, Peaks were given highest priority. They were moved the least, deleted the least, and most often recognized when missing. Peaks were also viewed far longer when moved to the front of a strip or out of order, while strips with deleted Peaks were also rated as the least coherent of all strips. These results are predicted by VNG: Peaks hold the most salient information about the interactions and events of the sequence. Thus, if missing, greater inference would be needed to deduce meaning, while the impact of that missing information would be felt the strongest.

6.1.2. Initials

While Initials also appeared important to the sequence, they were far less consistent than Peaks. In many cases, their distributions fell in between the "peripheral" categories and Peaks. They were among the least deleted, most accurately recognized when missing, and least proportionally moved in reconstruction. However, these rates often did not differ significantly from chance. Also, Initials were inconsistent in where they were moved within a sequence, except toward the beginning of a strip—reflecting their status as "initiating" panels. Indeed, Initials at the start of a sequence still were viewed as fast as Establishers in Experiment 2, yet when Initials ended a strip, they were viewed the longest, and their sequences were rated the least coherent of all sequences.

6.1.3. Initiating peaks

Initiating Peaks reflected traits of both Initials and Peaks in distribution. Like Initials and Peaks, they were deleted in low proportions, and were often recognized when missing. However, in reconstruction, Initiating Peaks were misplaced far more often than chance, unlike Peaks. Yet unlike Initials, which were fronted, Initiating Peaks were misplaced in Position 3—the canonical Peak position—six times more than to any other position. Finally, ratings of coherence of strips with missing Initiating Peaks also fell directly between ratings for Initials and Peaks.

Given these findings, Initiating Peaks provide evidence for the distributional assignment of categorical roles. Visual Narrative Grammar proposes that Initiating Peaks are a particular kind of Peak that can motivate a second role as an Initial at a higher level of constituent structure. While these results cannot support the presence of constituent structure, these distributional tasks do suggest that Initiating Peak panels can at least play multiple roles. If these panels were coded ambiguously, we would have observed qualities similar to Initials and Peaks, but clearly not aligning with just one. With no overarching theory to guide us, we might posit they were a unique category with overlapping traits of Initials and Peaks. However, because the theory proposes that these panels *should* be similar to both, it need not be labeled as a unique category, but as a hybrid with semantic features allowing it to play multiple roles.

6.1.4. Releases

While Releases seem to sometimes be peripheral, they trended toward having some importance to the sequence. For example, Releases were frequently deleted, and often not recognized when missing. However, despite being the *most* misplaced category in non-canonical strips (along with Prolongations), they were the *least* misplaced in canonical strips. Strips with missing Releases were also rated among the least coherent of all strips (Experiment 4), as were strips with Releases moved forward in the sequence (Experiment 2). In sum, although they often appear somewhat peripheral (as in deletions and recognitions), Releases have contexts where they are more important to the sequence.

One reason for the difference between Releases and other peripheral categories may be their position in the sequence. Strips with deleted panels toward the front of a sequence (Establishers, Initials, and Prolongations) were rated as more coherent than those with deleted panels toward the end of the sequence (Peaks, Releases). This echoes Lakusta and Landau's (2005) observation of the saliency of goals over sources in the structure of paths in the language of children and adults. Regier (1996, 1997) also suggested a preference for endpoints over starting points of paths in perception and attention. Such a preference would support why Releases may be less important overall, but more noticeable when missing or moved out of order.

6.1.5. Establishers

Establishers appeared to be clearly peripheral to the narrative. Establishers were deleted the most often, recognized when missing the least often, and strips with missing Establishers were rated the most coherent no matter the strip type—their absence made little difference in comprehension. They were also viewed the fastest of all panels at both the front and end of sequences. This was expected: Establishers provide "set up" information about a sequence available in other panels. This makes Establishers easy to delete and is consistent with the previous research showing that beginnings of verbal and filmed narratives are expendable (Kraft et al., 1991; Mandler, 1984; Mandler & Johnson, 1977).

If they are so unimportant, what good are these panels? Narratively, Establishers devote a panel to establishing information without acting upon it, as opposed to introducing characters along with their actions. This would facilitate what Gernsbacher (1990) has described as the process of "laying a foundation" with new information relevant for a discourse. Since readers likely use such a process at the start of a sequence anyhow (as reflected in the slower viewing times at Position 1 in Experiment 2), the narrative grammar provides this function to a distinct narrative unit.

6.1.6. Prolongations

Prolongations were also clearly less important to the sequence than other categories. They were moved the most often, guessed infrequently when missing, and strips without them were rated as very coherent. These results were expected given that Prolongations function as a medial state between an Initial and Peak. They act as a "buffer" between the more important information of a sequence, perhaps for pacing or suspense. In terms of meaning, Prolongations are largely unneeded outside of specifying the manner of a path action or providing a pause in timing of a depiction (or, more superfluously, serving as "filler" when a strip's meaning only requires three panels, but an author must fill four). Prolongations serve a largely narrative—as opposed to semantic—function, and thus are less essential for the meaning of a strip but can enrich how it is conveyed.

6.2. Distributional trends

While distributional tendencies arose for individual categories, similarities appeared between categories as well. In broad strokes, Establishers, Prolongations, and Releases were deleted at higher proportions and recognized at lower proportions. Also, these categories were frequently misplaced in positions associated with each other. Establishers moved to the ends of sequences (to function as Releases). Establishers and Releases moved between Initials and Peaks (to function as Prolongations). Also, Releases and Prolongations frequently moved to the start of a sequence (to function as Establishers). Furthermore, in Experiment 2, Establishers and Releases were viewed at comparable paces at both the starts and ends of sequences (though, contrastingly, when reversed were rated as fairly incoherent).

Overall, these trends imply distributional functionality between these categories. There are two ways to interpret this:

1. *Single category*: Similarities in distributional tendencies tell us that these panels are all actually a single formal category that happens to functionally appear in different places at different times, such as the way Nouns functionally appear in both Subject and Object positions.
2. *Functional context*: The differences between distributional tendencies are great enough to support that they are actually different categories. These categories "receive" panels that have similar semantic representations, but assign them different functional roles in the global narrative sequence.

These approaches have nuanced differences. The "Single Category" approach defines categories on the fixed basis of properties of the individual panel, which can then appear in different positions in the strip. The "Functional Context" approach involves the interaction between the properties of the panel and their role in the global sequence. VNG argues that this Functional Context viewpoint is more feasible, given that narrative structure does not mandate a one-to-one mapping with semantics (Cohn, 2013; Cohn et al., 2012). Establishers, Prolongations, and Releases are all functional categories in the global sequence that may allow for panels with similar semantic representations of "passive states" to take those roles. Each category plays a unique role in the scope of the narrative sequence, but it may use panels with similar content to do so.

This viewpoint helps to clarify the contrast between core and peripheral panels. Prototypically, Initials and Peaks are more robust in their semantic representations: They show the central parts of events and thereby directly depict causal relations. In contrast, Establishers, Prolongations, and Releases typically depict more passive states and appear more flexible and expendable in distribution than the core categories. This difference means that "core" categories are more deterministic in their mapping of semantics to narrative, while "peripheral" categories are more flexible. This provides at least limited support for a weak version of the idea that discourse units can play varying roles in a sequence depending on their context (Jahn, 1997; Sternberg, 1982), although this flexibility is contingent upon the kinds of information a panel contains. We might imagine then, in contrast to the passive states of prototypical Establishers and Releases, that culminating completed actions could feasibly act as Establishers in highly constrained contexts. However, this "non-prototypical" mapping of semantics to narrative may incur a cost in processing (suggested by the slower viewing times to Peaks in first position in Experiment 2). Thus, the structure of sequential images involves an interaction between the semantic characteristics of individual panels and corresponding narrative roles in a global schema.

Several questions can follow these results: How do these narrative categories behave in longer and more complicated sequences? How do these categories impact the online processing of sequential images? Can these categories be extended to the study of verbal discourse, and do verbal narrative units involve similar semantic cues? These questions can help motivate the future study of narrative structure in the visual and verbal form.

Acknowledgments

Phillip Holcomb is thanked for funding this work through the grant NIH HD25889; and support was also provided by the Tufts Center for Cognitive Studies. Previous drafts greatly benefited from feedback from Ray Jackendoff. Tali Ditman, Brian McCloone, Martina Santarsieri, and Ben Horn aided in data analysis. Fantagraphics Books generously donated *The Complete Peanuts*.

Note

1. It is important to note that any narrative category can function as a whole phase, not just Initials. Peak panels are always the "heads" of these constituents. Thus, a "Releasing Peak" might be a panel that plays a Peak locally but provides the head of a Release phase. All narrative categories have this recursive capacity, described further in Cohn (2013). In this experiment, we focus primarily on the "dual" role of Initiating Peaks.

References

Abel, J., & Madden, M. (2008). *Drawing words and writing pictures*. New York: First Second Books.

Arijon, D. (1976). *Grammar of the film language*. London: Focal Press.

Black, J. B., & Bower, G. (1980). Story understanding as problem-solving. *Poetics*, *9*, 223–250.

Bordwell, D. (2007). *Poetics of cinema*. New York: Routledge.

Bordwell, D., & Thompson, K. (1997). *Film art: An introduction* (5th ed.). New York: McGraw-Hill.

Bresman, J. (2004). *Test of sequential "sense" of comics*. Unpublished class essay: MSTU 5510.008: Social and Communicative Aspects of the Internet and Other ICTs, Teacher's College, Columbia University, New York.

Brewer, W. F., & Lichtenstein, E. H. (1981). Event schemas, story schemas, and story grammars. In J. Long & A. D. Baddeley (Eds.), *Attention and performance IX* (pp. 363–379). Hillsdale, NJ: Erlbaum.

van den Broek, P. (1994). *Comprehension and memory of narrative texts: Inferences and coherence*. New York: Academic Press.

Butcher, S. H. (1902). *The poetics of aristotle* (3rd ed.). London: Macmillian and Co., Ltd.

Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.

Cohn, N. (2003). *Early writings on visual language*. Carlsbad, CA: Emaki Productions.

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

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.

Cohn, N., Paczynski, M., Jackendoff, R., Holcomb, P. J., & Kuperberg, G. R. (2012). (Pea)nuts and bolts of visual narrative: Structure and meaning in sequential image comprehension. *Cognitive Psychology*, *65*(1), 1–38.

Freytag, G. (1894). *Technique of the drama*. Chicago, IL: S.C. Griggs & Company.

Gernsbacher, M. A. (1983). *Memory for surface information in non-verbal stories: Parallels and insights to language processes*. Unpublished dissertation, University of Texas at Austin, Austin, TX.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Lawrence Earlbaum.

Haberlandt, K. (1980). Story grammar and reading time of story constituents. *Poetics*, *9*(1–3), 99–118.

Haberlandt, K., Berian, C., & Sandson, J. (1980). The episode schema in story processing. *Journal of Verbal Learning and Verbal Behavior*, *19*(6), 635–650.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Halliday, M. A. K., & Hasan, R. (1985). *Language, context, and text: Aspects of language in a social-semiotic perspective*. Victoria: Deakin University Press.

Hinds, J. (1976). *Aspects of Japanese discourse*. Tokyo: Kaitakusha Co., Ltd.

Hobbs, J. R. (1985). *On the coherence and structure of discourse*. Stanford, CA: CSLI Technical Report 85-37.

Huber, W., & Gleber, J. (1982). Linguistic and nonlinguistic processing of narratives in aphasia. *Brain and Language*, *16*, 1–18.

Jackendoff, R. (1990). *Semantic structures*. Cambridge, MA: MIT Press.

Jahn, M. (1997). Frames, preferences, and the reading of third-person narratives: Towards a cognitive narratology. *Poetics Today*, *18*(4), 441–468.

Kaufman, A. S., & Lichtenberger, E. O. (2006). *Assessing adolescent and adult intelligence* (3rd ed.). Hoboken, NJ: Wiley.

Keenan, J. M., Baillet, S. D., & Brown, P. (1984). The effect of causal cohesion on comprehension and memory. *Journal of Verbal Learning and Verbal Behavior*, *23*, 115–126.

Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. Stanford, CA: CSLI Publications.

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

Kunzle, D. (1973). *The history of the comic strip* (Vol. 1). Berkeley: University of California Press.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potential reflect semantic incongruity. *Science*, *207*, 203–205.

Lakusta, L., & Landau, B. (2005). Starting at the end: The importance of goals in spatial language. *Cognition*, *96*, 1–33.

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

Magliano, J. P., & Zacks, J. M. (2011). The impact of continuity editing in narrative film on event segmentation. *Cognitive Science*, *35*(8), 1489–1517.

Mandler, J. M. (1978). A code in the node: The use of story schema in retrieval. *Discourse Processes*, *1*, 14–35.

Mandler, J. M. (1984). *Stories, scripts, and scenes: Aspects of schema theory*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Mandler, J. M., & DeForest, M. (1979). Is there more than one way to recall a story? *Child Development*, *50*, 886–889.

Mandler, J. M., & Goodman, M. S. (1982). On the psychological validity of story structure. *Journal of Verbal Learning and Verbal Behavior*, *21*, 507–523.

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, *9*, 111–151.

Mann, W. C., & Thompson, S. A. (1987). *Rhetorical structure theory: A theory of text organization*. (No. ISI/RS-87-190). University of Southern California, Marina Del Rey Information Sciences Institute.

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

McCloud, S. (2006). *Making comics*. New York: Harper-Collins.

Myers, J. L., Shinjo, M., & Duffy, S. A. (1987). Degree of causal relatedness and memory. *Journal of Memory and Language*, *26*(4), 453–465.

Nakazawa, J. (2002). Analysis of manga (comic) reading processes: Manga literacy and eye movement during Manga reading. *Manga Studies*, *5*, 39–49.

Nakazawa, J. (2004). Manga (comic) literacy skills as determinant factors of manga story comprehension. *Manga Studies*, *5*, 7–25.

Nakazawa, J., & Nakazawa, S. (1993). How do children understand comics?: Analysis of comic reading comprehension. *Annual of Research in Early Childhood*, *15*, 35–39.

Nezworski, T., Stein, N. L., & Trabasso, T. (1982). Story structure versus content in children's recall. *Journal of Verbal Learning and Verbal Behavior*, *21*(2), 196–206.

Regier, T. (1996). *The human semantic potential: Spatial language and constrained connectionism*. Cambridge, MA: MIT Press.

Regier, T. (1997). Constraints on the learning of spatial terms: A computational investigation. In R. Goldstone, P. Schyns, & D. Medin (Eds.), *Psychology of learning and motivation: Mechanisms of perceptual learning* (Vol. 36, pp. 171–217). San Diego, CA: Academic Press.

Rumelhart, D. E. (1975). Notes on a schema for stories. In D. Bobrow & A. Collins (Eds.), *Representation and understanding* (pp. 211–236). New York: Academic Press.

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

Saraceni, M. (2003). *The language of comics*. New York: Routeledge.

Schulz, C. M. (2004a). *The complete Peanuts: 1950–1952*. Seattle, WA: Fantagraphics Books.

Schulz, C. M. (2004b). *The complete Peanuts: 1953–1954*. Seattle, WA: Fantagraphics Books.

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.

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), *New directions in discourse processing* (pp. 53–119). Norwood, NJ: Ablex.

Stein, N. L., & Nezworski, T. (1978). The effects of organization and instructional set on story memory. *Discourse Processes*, *1*(2), 177–193.

Sternberg, M. (1982). Proteus in Quotation-Land: Mimesis and the forms of reported discourse. *Poetics Today*, *3*(2), 107–156.

Talmy, L. (2000). *Toward a cognitive semantics* (Vol. 1). Cambridge, MA: MIT Press.

Thorndyke, P. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, *9*, 77–110.

Trabasso, T., Secco, T., & van den Broek, P. (1984). Causal cohesion and story coherence. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 83–111). Hillsdale, NJ: Erlbaum.

Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, *24*, 612–630.

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

Zwaan, R. A., & Rapp, D. (2006). Discourse comprehension. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholingusitics* (2nd ed., pp. 641–674). New York: Academic Press.