From Visual Narrative Grammar to Filmic Narrative Grammar:

The narrative structure of static and moving images

Neil Cohn

Tilburg University

www.visuallanguagelab.com

© 2016. Neil Cohn.

Pre-proof version. For final version, see:

Cohn, Neil. 2016. "From Visual Narrative Grammar to Filmic Narrative Grammar: The narrative structure of static and moving images." In *Film text analysis: New perspectives on the analysis of filmic meaning*., edited by Janina Wildfeuer and John A. Bateman, 94-117. London: Routledge.

Introduction

The question of how we understand sequences of communicative images applies equally to static sequential images—as in the visual narratives found in comics—and to moving sequential images—as in the visual narratives found in film. Many assume that similar underlying principles of comprehension apply across both static and moving images, though they present information in ways that can depart radically. This paper explores the comprehension of sequential images by exploring how the theory of *Visual Narrative Grammar* (VNG) (Cohn 2013b) overlaps with notions from film theories and can apply to films. Since VNG has been designed as a theory for static sequential images (like those found in comics), we will in particular explore how moving images change this structure to suit the affordances of the medium, adapting to become a *Filmic Narrative Grammar* (FNG).

The challenge at hand

Drawn sequential visual narratives have appeared throughout human history (Kunzle 1973, McCloud 1993), yet until the advent of film, the illusion of motion had been fairly rudimentary and limited in scope. Despite this longevity for drawn narratives, theories about the structure of films have been far more prevalent and have a longer contiguous history than most theories of static visual narrative comprehension. In both cases, early theories appealed to unitary inferential processes, like "montage" (Eisenstein 1942) or "closure" (McCloud 1993) for almost all relations between images. More recent theories from psychology have stressed that film comprehension, and the development of its editing, adapts the preferences of general cognitive processes related to attention, perception, and event segmentation (Berliner and Cohen 2011, Levin and Simons 2000, Smith 2012, Zacks 2014). These theories link well with psychological approaches to discourse and narrative which stress the monitoring of linear semantic relationships throughout a visual narrative (Magliano and Zacks 2011, Zacks 2014, Zwaan and Radvansky 1998), and echo the classic focus on the inferential processes fusing together images or film shots (Eisenstein 1942, McCloud 1993). Such models are also consistent with theoretical approaches to visual narratives that emphasize the dynamic updating of a mental model with semantic information (e.g., Bateman and Wildfeuer 2014, Wildfeuer 2013).

Interestingly, these same basic constructs are posited as not only guiding our understanding of visual narratives, but also our understanding of daily life (Radvansky and Zacks 2014, Zacks 2014). Given that all individuals should share basic perceptuo-semantic cognitive processes, such understandings should therefore be universal and uniform. However, why then are some individuals who lack of exposure to film unable to understand certain sequences (Ildirar and Schwan 2015, Schwan and Ildirar 2010)? And, why are some individuals are unable to comprehend or produce coherent drawn sequential images without experience with graphic narratives (e.g., Byram and Garforth 1980, Fussell and Haaland 1978, Wilson 2016), even with exposure to film (Wilson and Wilson 1987)? Rather, exposure and practice with a graphic system matters, evident even in neural responses to narrative patterns, which are modulated by readership of specific types of visual narratives that use those patterns (Cohn and Kutas In prep).

In addition, while inference and general perception may help explain aspects of sequential image *comprehension*, they do little to explain *production*. That is, they cannot explain the processes governing how an author selects which image comes *next*, especially when such sequences are drawn, not "edited" together. A process of dynamic updating and/or perceptuo-semantic inference only explains a *backward-looking* process integrating what has

already been experienced, but does not constrain what might be produced or comprehended *ahead* of what has already been read or drawn. Yet, comprehenders *do* make such forward-looking predictions in visual narratives (Cohn et al. 2014, Cohn and Paczynski 2013).

Furthermore, growing evidence supports that comprehenders do not rely solely on changes in meaning between images in a sequence (Cohn et al. 2014, Cohn et al. 2012, Cohn and Bender Under Review). Experimentation has long shown that images form groupings into constituents beyond linear relations (Gernsbacher 1985, Hagmann and Cohn 2016), and such groupings rely on forward-looking predictive processes, not just backward-looking reanalysis of prior relations (Cohn et al. 2014). Sequences may also have structural features that cannot be processed linearly, such as long-distance connections between images, center-embedded clauses, and structurally ambiguous sequences with several coherent parsings (Cohn 2013b, 2010). And again, such structures are not dependent on semantic relations (Cohn and Bender Under Review, Cohn et al. 2014, Cohn and Kutas In prep).

Altogether, these issues warrant a theory of sequential image structure that extends beyond inference and general perceptuo-semantic processes alone, as in the emergent psychological literature on film. At the same time, theoretical accounts of visual narrative structure must offer more than just a promissory note as to the underlying cognitive processes, by providing testable predictions that are in turn supported by the psychological literature.

Visual Narrative Grammar

Visual Narrative Grammar (VNG) outlines a model of sequential image understanding integrating linguistic theory and psychological experimentation (Cohn 2013a). In that it draws an analogy between language and sequential images, VNG is similar to previous "linguistic" approaches to film (Metz 1974), including those that have attempted to model filmic "grammar" (Carroll 1980, Buckland 2000, Colin 1995, Chateau 1987). However, unlike previous theories, VNG is based on contemporary linguistic theories that separate grammar from the meaning (Culicover and Jackendoff 2005, Jackendoff 2002), rather than on outdated "structuralist" theories (e.g., Metz 1974) or Chomskyan phrase structure grammars (Chomsky 1965)—both of which have ambiguous relations between structure and meaning (e.g., Carroll 1980, Buckland 2000, Colin 1995) and do not account for decades of psychological research showing such a separation (e.g., Marslen-Wilson 1987, Osterhout and Nicol 1999, Miller and Isard 1963).

In addition, unlike many theoretical models, VNG's claims are grounded in linguistic *methods* like diagnostic tests (Cohn 2013b, 2014a) that form the basis for psychological experimentation (Cohn 2014b, Cohn and Bender Under Review, Cohn and Wittenberg 2015). These studies have shown that similar neural responses appear to manipulations of VNG as to manipulations of syntax in sentences, and that these are different from the neural responses to semantics (Cohn et al. 2014, Cohn et al. 2012). VNG is thus distinguished not only from older "grammatical" models, but also from models of sequential image understanding that posit only meaningful relations between units (Bateman and Wildfeuer 2014, Magliano and Zacks 2011). Such evidence grounds VNG in empirical, psychological research that extends beyond just theoretical claims.

VNG argues that semantic cues within images in a sequence map to narrative roles organized within hierarchic constituents at a "discourse" level of meaning, analogous to the way that syntactic categories organize words into constituents at a sentence level (Cohn 2013b). Though VNG has been designed to describe the structure of drawn sequential images (as in comics), I argue here that it can also apply to filmic narratives, which will be described in depth.

VNG posits several basic narrative categories[1]:

Establisher (E) – sets up an interaction without acting upon it, often as a passive state
Initial (I) – initiates the tension of the narrative arc, prototypically a preparatory action and/or a source of a path
Peak (P) – marks the height of narrative tension and point of maximal event structure, prototypically a completed action and/or goal of a path, but also often an interrupted action
Release (R) – releases the tension of the interaction, prototypically the coda or aftermath of an action

These descriptions outline the prototypical correspondences of narrative categories to meaning—the ways that the *semantic* content (the visual cues within images) may influence the *structural* role that an image plays in a sequence. Nevertheless, narrative categories in VNG are identified both by a panel's bottom-up content and its top-down context in a global sequence, as determined by distributional tendencies throughout a narrative sequence (Cohn 2013b, 2014b). This is analogous to how syntactic categories (like nouns, verbs) prototypically correspond to the semantics (like objects, events) of words (Jackendoff 1990), but also rely on sentence context.

The canonical narrative schema in VNG places these categories in this order:

Canonical narrative schema
[$_{\text{Phase X}}$ (Establisher) – (Initial) – Peak – (Release)]

This schema outlines that a "phase" (a narrative constituent) consists of these categories in this order. Not all phases need to contain all categories, and thus the non-obligatory elements are in parentheses. Because Peaks motivate a sequence (as the "head" of the phase), it is not marked as optional. However, even Peaks can be omitted felicitously under certain constrained, inference-generating conditions (Cohn and Kutas 2015, Magliano et al. 2015). Note that this narrative schema is not a "rule" in the sense of traditional phrase structure grammars (e.g., Chomsky 1965 for syntax, Mandler and Johnson 1977 for narrative, Carroll 1980 for film), but rather is a "construction" stored in memory as an abstract schematic pattern, akin to syntactic patterns in the lexicon of language (Goldberg 1995, Jackendoff 2002).

Narrative categories do not just apply to individual images, but each category can expand into its own phase. Consider the sequence in Figure 1. In the first panel, a boxer reaches back to punch another, which is a preparatory action prototypical of an Initial. The punch is then completed in the next panel, a Peak in relation to that Initial. The next panel resets the actions with an Establisher in panel 3, which sets up a new situation in a relatively passive action of the boxers standing across from each other. Another Initial appears in panel 4, again with a preparatory action. The next panel "zooms in" on the information in the prior Initial with a spatial modifier of a ***Refiner*** (discussed below). The penultimate panel, a Peak, does not depict a completed action, but rather shows an unexpected interruption of the boxer's action: slipping and making the punch unrealized. The final panel shows the coda of the action in the prior Peak, here with a Release showing the victor standing over his knocked-out opponent.

[1] Note: some categories such as Prolongations and Orienters are omitted for simplicity (Cohn 2013b).

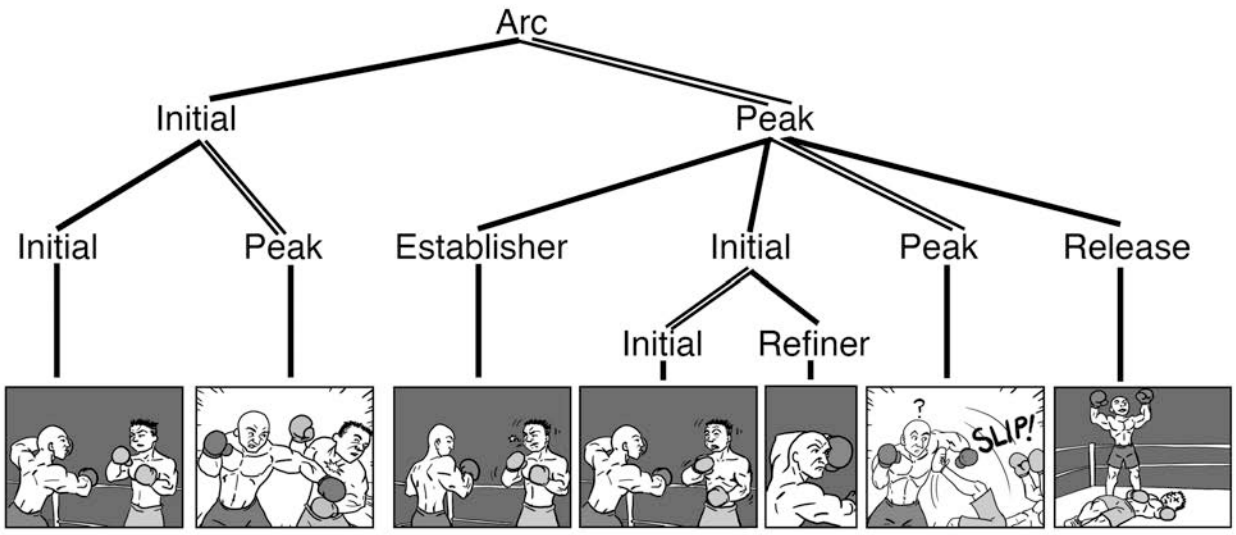

Figure 1. A narrative sequence with two narrative constituents and one subordinate modifying constituent with a Refiner.

Beyond the roles that panels play, this sequence also embeds panels into hierarchic structures. As is, the surface pattern of *I-P-E-I-Ref-P-R* does not conform to the canonical narrative schema. However, its parts do, if combined to form constituents. Thus, complexity can primarily be introduced into a narrative by expanding categories into their own constituents. Any grouping of panels can play a role in a larger structure, and when they reach a maximal node, the whole is considered an "Arc." In Figure 1, the first two panels form an Initial constituent that together set up a Peak constituent comprised of the remaining five panels. Each of these constituents maintains the canonical narrative schema within their scope. As discussed before, Peaks form the "head" of each phase, now indicated by double-barred lines to show that they motivate the primary meaning of their superordinate phase (i.e., each constituent is an expansion of its Peak). Thus, narrative categories apply recursively to individual panels and whole constituents. These principles also extend further upward—the same principles that guide short sequences govern higher "plot" level narrative structures.

VNG expands on the canonical narrative arc using several modifiers, such as Refiners, which repeat information found in another panel to give it a "refined" viewpoint (Cohn 2013a, 2015b). In Figure 1, the second Initial expands into a phase when modified by a Refiner of the puncher alone. Here, the larger viewpoint panel becomes the motivating "head" panel (double bar lines) and the Refiner is its modifier. Any category (X) can thus be expanded with modifiers (Refiners) either before or after a head, within a phase of that same category (Phase X).

Head-Modifer Schema
[$_{\text{Phase X}}$ (Modifier) – X – (Modifier)]

Perspective Shifts follow the same head-modifier relationship as Refiners, but instead of honing in on information, they provide a less prototypical spatial viewpoint of the same information. For example, if Figure 1 replaced the Refiner with an aerial viewpoint of the same action as in the Initial panel (boxer reaching back to punch the other), this panel would be a Perspective Shift on the prototypical viewpoint (a lateral view) in the prior panel. Both Refiners and Perspective Shifts are "modifiers" in the Head-Modifier Schema above. If a head were to be

deleted, the modifier takes the role of its head: if the Initial in Figure 1 were deleted, the Refiner would become an Initial, without needing a larger phase.

The other primary way to expand a sequence repeats narrative categories (Cohn 2015b). ***Conjunction*** occurs when a single category (X) repeats within a phase of that category (Phase X):

Conjunction schema
$[_{\text{Phase X}}\ X_1 - X_2 - \ldots\ X_n]$

This notion of "conjunction" is different from the semantic relations often posited in discourse and film theories (Bateman and Schmidt 2012, van Leeuwen 1991, Martin 1983), here aligning with conjunction from syntactic theory: successive panels play the same grammatical role within a constituent of that category. These units can then map to meaning in several ways (Cohn 2015b). Figure 2 depicts several potential semantic correspondences to an Initial constituent containing three conjoined panels. Along the lower left tier, each three-panel sequence shows an option for an Initial sequence showing (a) actions or events (A-Conjunction), (b) characters within a scene (E-Conjunction), (c) parts of a single character (N-Conjunction), or (d) disparate semantically associated elements (S-Conjunction). These conjoined images at left create the conceptual equivalent of the panels to the right. That is, the three conjoined images could be replaced using a single, non-conjoined image in the right tier (a diagnostic for assessing conjoined panels). Thus, Figure 2 depicts different types of *semantic* manifestations of the same *narrative* conjunction.[2]

[2] Note that the framing categories listed on the right of Figure 2 (macro, mono) do not precisely map to "shot types" in film. While a prototypical macro may use a long shot, macros are defined simply by the interaction of multiple characters in a frame. Mono panels likewise depict only a single character, whether shown in a long, full, or close shot (Cohn 2013a). Thus, these categories are about quantity of information, not how that information is presented.

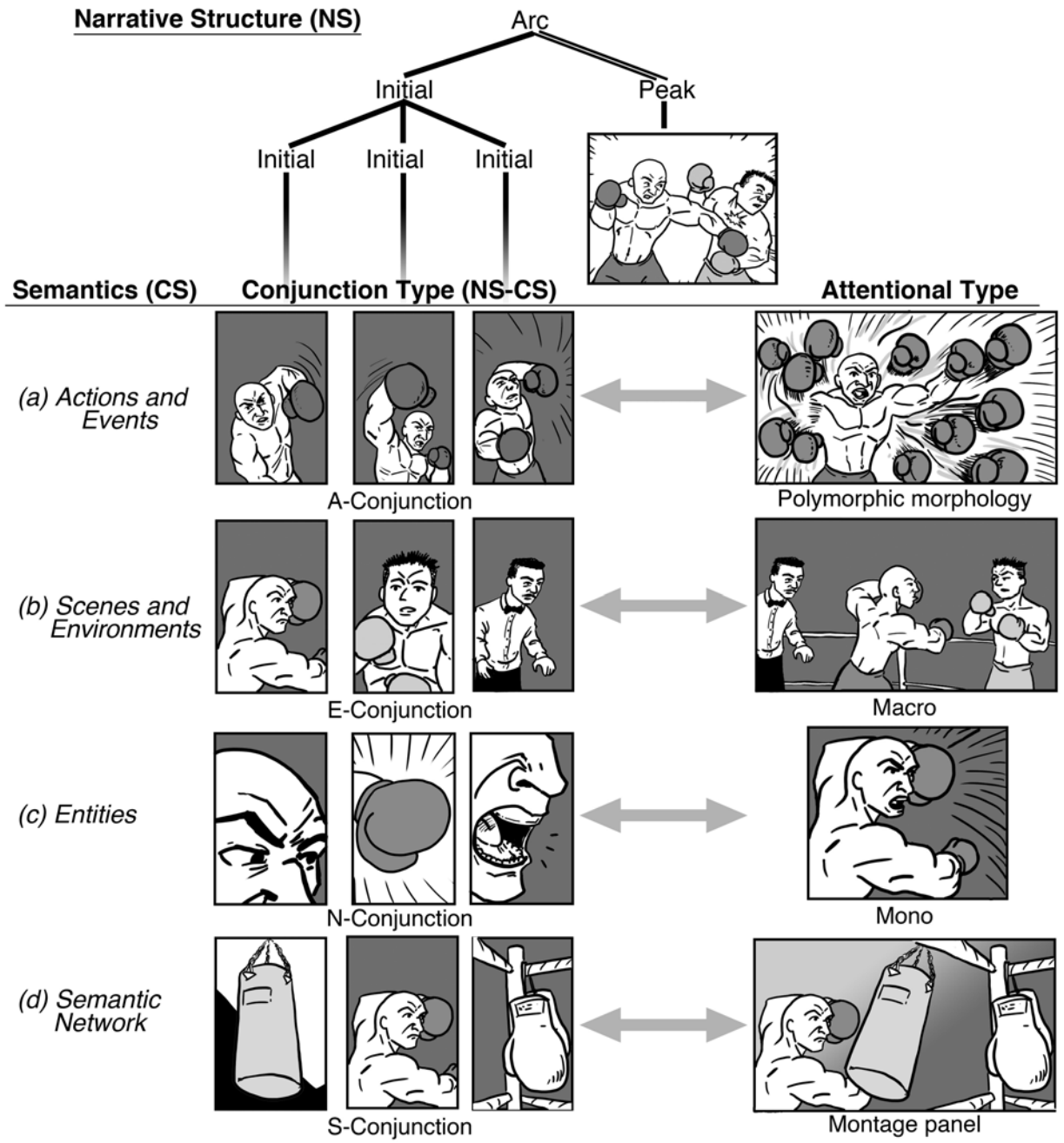

Figure 2. Different types of narrative conjunction using the repetition of a single narrative category (Initials) to show various semantic information (actions, characters in a scene, parts of an individual, or semantically associated elements), which could also be framed by a single image.

Altogether, these three narrative schemas, stored in long-term memory as "constructions", align with the same basic principles of combination found at the syntactic level (Culicover and Jackendoff 2005, Jackendoff 2002). However, human languages do not just use basic abstract combinatorial schemas—there are thousands of constructions that use the basic schemas in regularized ways or depart from those canonical patterns. Similarly, VNG allows for

constructional patterns beyond these basic schemas, both using and departing from these basic principles.

VNG and film theories

I have now outlined the basic constructs of VNG as they have been proposed for static visual narratives. Among the characteristics of this model are categorical roles played by units and constituents, hierarchic structures that allow connections across distances, modifiers that expand on basic sequencing, and the storage of these elements as constructional patterns in memory. We can now ask: How might these structures relate to those posited in film theories?

Constructional patterns in film

Let's start by looking at Metz's (1974) *grande syntagmatique* for the "language" of film. Metz outlines several patterned ways that film shots relate to each other, such as semantic associative relations, alternating shots, and temporally progressive but non-narrative relations, among others. Though posited as articulating syntagamatic relations—i.e., how shots are strung into sequences—they actually reflect a more paradigmatic character—i.e., patterned regularities that subsequently appear in sequences (Bateman 2007, Bateman and Schmidt 2012). Subsequently, Metz's approach has been dissected and reinterpreted as motivated by underlying primitives which can better characterize a range of relationships between shots (Bateman 2007, Bateman and Schmidt 2012).

On the one hand Metzian (and subsequent) relations *do* characterize connections that extend across sequential shots (syntagmatic), but on the other, they are systematic conventions that comprise specific regularities (paradigmatic), not abstract combinatorial "rules." Such tension is precisely what we would expect in *grammatical constructions* (Goldberg 1995, Jackendoff 2002)—they are sequential patterns that are instantiated in the lexicon of a language. They are at once grammatical (syntagmatic) and lexical (paradigmatic).

Many of the Metzian patterns are incorporated as constructions within VNG, often emerging from mappings between the abstract combinatorial rules and varying semantic meanings. Consider the varying ways that conjunction maps to semantics in Figure 2. Metz's "episodic syntagma" are similar to A-Conjunction, while "bracket syntagma" are similar to S-Conjunction (among others). In VNG, these "types" are recognized as unique patterns, yet all use a common underlying combinatorial principle of conjunction (Cohn 2015b). VNG also provides such constructions with information beyond just the characterization of patterned semantic relations—such patterns play functional roles in the context of a sequence (discussed below). Because VNG is a construction grammar, it also allows novel idiomatic patterns, be they included in Metz's taxonomy or others (e.g., Bateman 2007, Bateman and Schmidt 2012, Branigan 1992).

Because VNG combines constructions in complex ways, other surface patterns may emerge from underlying combinatorial structure. Take for example Metz's alternating syntagma, also known as "crosscutting" or "multitracking" (Bateman and Schmidt 2012, Bordwell and Thompson 1997), where images flip back and forth between the actions of different characters. Consider Figure 3, from the original *Star Wars* (1977), which shows R2-D2 getting captured by droid-selling Jawas.

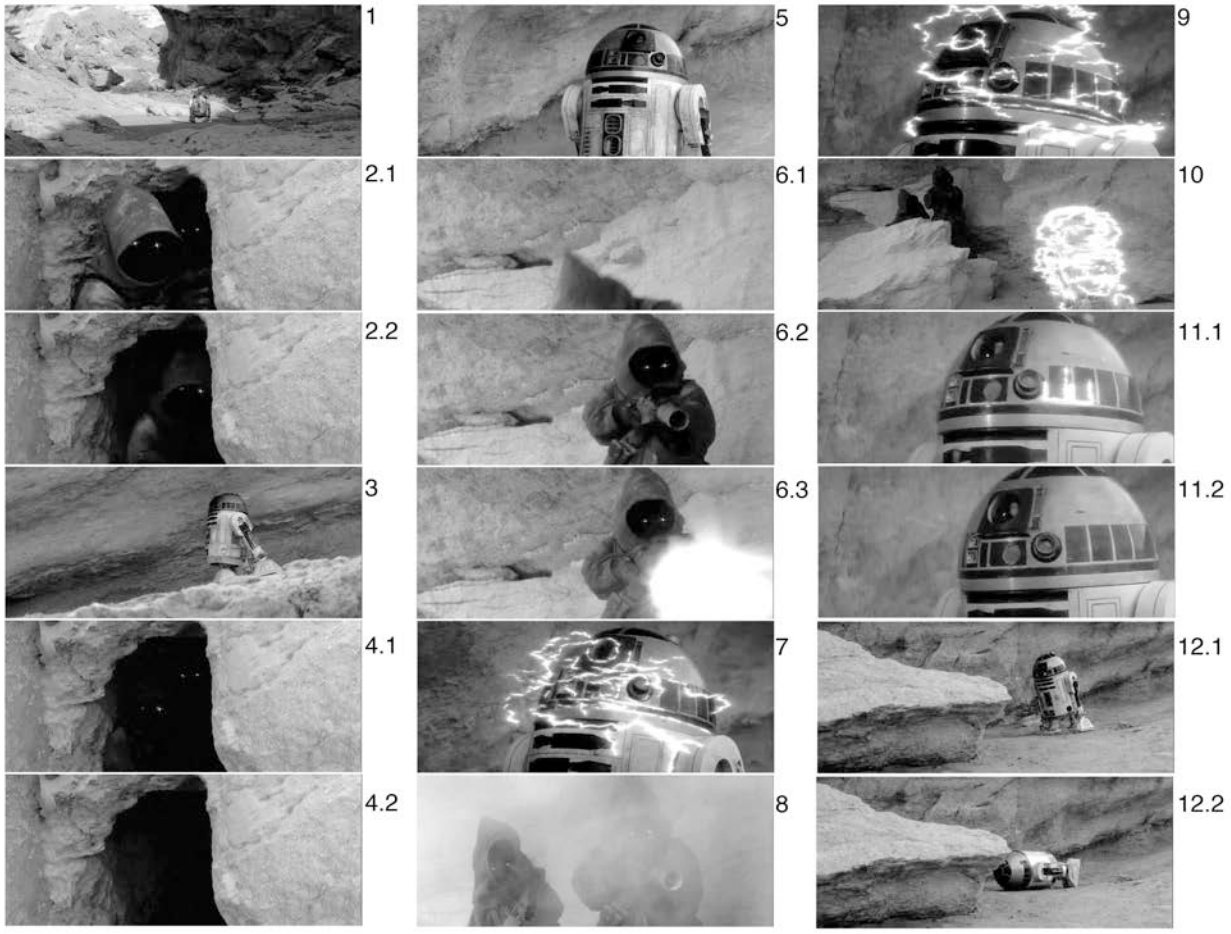

Figure 3. Narrative grammar applied to a sequence from *Star Wars* (1977: 12:00-12:27)

Analyzed by VNG, this particular sequence opens with Establishers, which show R2-D2 rolling through a rocky landscape (shot 1), then shows Jawas watching (2.1) him as they back into the shadows (2.2). These E-Conjoined Establishers set up the characters' relationship; yet, because a distance separates them, the characters occupy their own panels, and the phase maps to

the inference of the spatial relationship between them (subscript "e").[3] As R2-D2 continues rolling (3) the Jawas watch him, (4.1) and dart off (4.2). These E-Conjoined Initials initiate the subsequent conjoined Peaks of R2-D2 (5) getting shot by the Jawas (6.1-3). Conjoined Establishers then set up the new relationship between the electricity striking R2-D2 (7) as the Jawas watch (8). This spreads around the droid (9) in a subsequent Initial, and reaches its climax as it surrounds R2-D2 (10) in the main Peak of the sequence, before the electricity dissipates in a Release (11.1). This causes R2-D2 to fall over lifelessly (12.1-2) in the final Release of the broader sequence.

This sequence is characterized by multitracking between R2-D2 and the Jawas. In VNG, each pair of shots constitutes an E-Conjoined constituent leading up to the main climax of the sequence. Though the surface form crosscuts between characters, functionally similar panels group together, potentially operating on different levels—i.e., the opening Establishers are hierarchically higher than the conjunctions embedded within the Initial constituent. In each case, the local relations between panels map to semantic structure (here not shown in full, but notated with "e") whereby a viewer infers that both characters are in the same spatial location, despite never seeing them together (Cohn 2015b). Notably, the climactic Peak provides the first shot where both Jawas and R2-D2 appear together. This alternation leading to a "convergence" is characteristic of many multitracked patterns (Cohn 2013a).

Given the complexity of this analysis, it is worth asking: How might we check if it is correct? Since VNG is committed to integration with empirical verification, analyses emerges through the use of diagnostic tests (deletion, manipulation, or substitution of units) which assess the categories and structure of sequences (see Cohn 2015a for a tutorial), and subsequently provide the basis for psychological experimentation (e.g., Cohn 2014b, Cohn et al. 2014, Hagmann and Cohn 2016). Here we apply one such test to confirm the roles broader narrative arc: a paraphrase. Because Peak panels motivate the top-level categories as the "head" of the constituent (i.e., other categories expand from the head Peak), we can drop out all non-Peaks from non-maximal nodes from Figure 3 (e.g., within the Initial and Peak constituents) to paraphrase the sequence. This results in Figure 4, which depicts these primary shots, providing an adequate paraphrase. Note that the alternation/convergence pattern still maintains, only now simplified to its core elements. Such a test provides evidence that these panels are playing particular functional roles in the sequence, and do so at different levels of structure.

[3] The "e" notation is shorthand. In actuality, the narrative structure for VNG is accompanied by additional components from a conceptual/event structure and a spatial structure, which specify these semantic relations explicitly. The "e" actually reflects the *interface* between the conceptual and spatial structures with those in the narrative structures (Cohn 2015b), here omitted for simplicity.

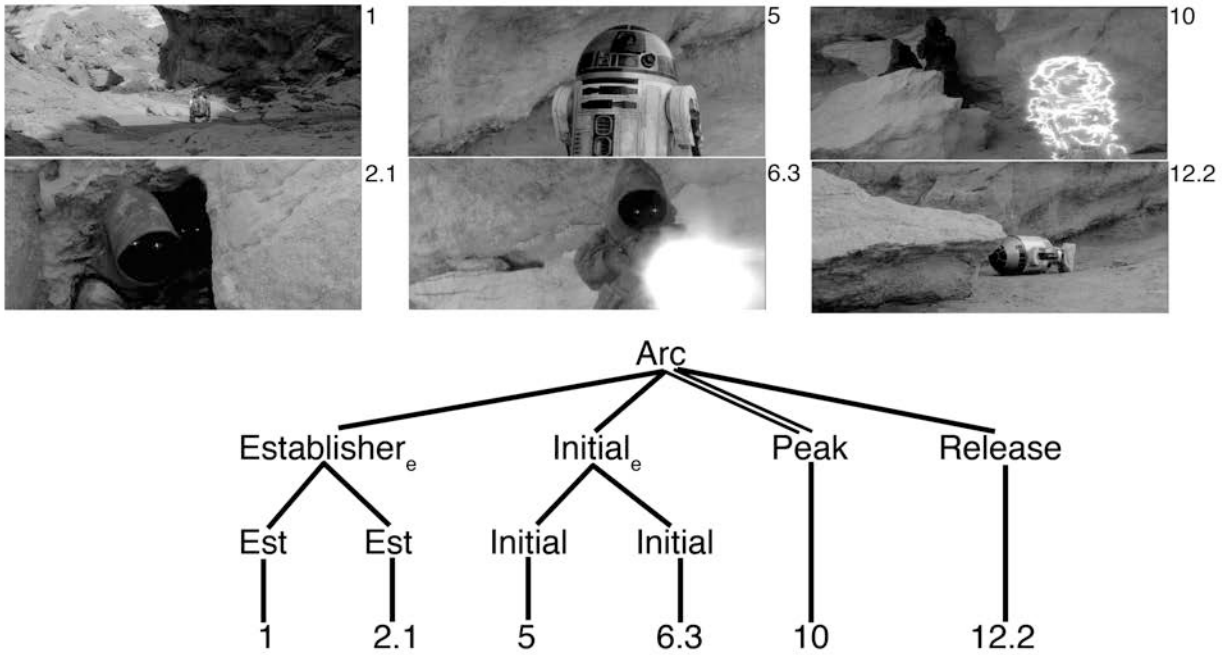

Figure 4. Paraphrase of the narrative grammar for the *Star Wars* (1977: 12:00-12:27) sequence in Figure 3.

Thus, in VNG, the Metzian alternation emerges from interactions between combinatorial structures, namely successive conjunctions. This interaction of structures may also be entrenched as a construction within the narrative grammar (Cohn 2013a). Note also that the successive conjunctions alone do not characterize this sequence: the top level maintains a canonical narrative arc, thereby allowing the lower-level shots and conjunctions to play categorical roles in relation to each other. This organization—facilitated by the parallel structures of narrative and meaning—allows the structure to avoid unnecessary transformational rules to manipulate one "deeper" sequence of separate tracks to become a "surface" linear sequence (Carroll 1980, Buckland 2000), or to require "shuffling" semantic structures into a layout (Bateman and Schmidt 2012).

If this "ABAB" alternation is indeed a construction stored in memory, then it should be preferable to others with the same component parts, as opposed to being an emergent pattern resulting from varying equal structural choices. In recent pilot research on drawn visual narratives, we used a forced choice test that asked participants to choose which pattern of sequence they prefer (N=33). Here, 78.5% of participants preferred ABAB alternation patterns to AABB patterns, while 90.5% preferred ABAB patterns to ABBA patterns. Such results imply that a preference is given to the alternation pattern over others—at least for drawn visual narratives—suggesting some evidence for a construction stored in memory.

In sum, like other approaches, VNG recognizes that alternating shots of character is a particular patterned structure (Bateman and Schmidt 2012, Metz 1974). Here, it is a surface pattern arising from underlying structures stored in memory as a construction (Cohn 2013a), and able to be modified (e.g., convergence, embedding). Such a construction uses interactions between structures for meaning (where the two tracks of events are comprehended) and basic narrative schemas that capture categorical roles of shots/panels relative to each other (where the linear sequence uses multitracking). In some ways this is analogous to the editing process:

multiple tracks of film footage (the unconstrained semantics) are edited together to form a linear sequence (narrative). However, in VNG, these separate structures are not external to a comprehender (as in editing), but rather are internal cognitive processes (Cohn et al. 2014, Cohn et al. 2012).

Hierarchic structure

As shown above, VNG is characterized by hierarchic structures. Experimental evidence of such hierarchies in film structure extend back decades (e.g., Carroll and Bever 1976), though more recently has combined both behavioral and neurocognitive methods (Zacks 2014). For example, participants consistently agree on where to segment short videos of events, and since fine-grained segmentations both align with and fall within coarse-grained segments, it implies a hierarchic organization (e.g., Zacks et al. 2001). In addition, brain activation from passive viewing of the same visual events correlates temporally with the boundaries between events later identified through a segmentation task (Zacks et al. 2001). Such boundaries have been speculated as being caused by the discontinuities in meaning between shots, such as changes in characters, location, or time (Magliano and Zacks 2011, Zacks et al. 2010). This idea is at least somewhat consistent with the structures found in semiotic film theories, which focus on the hierarchies created by dynamically changing aspects of meaning (Bateman and Schmidt 2012, Wildfeuer 2013).

In VNG, hierarchy is not dependent on meaning alone, and indeed recent research has shown that narrative categories (discussed below) provide a stronger predictor for participants' segmentations of a drawn sequence than do changes in meaning (Cohn and Bender Under Review). In addition, hierarchy in VNG is not uniform, but rather emerges from different schematic constructions (e.g., the canonical arc, conjunction, and the head-modifier schema), and their embedding within each other. For example, narrative categories recursively characterize individual panels/shots, whole constituents, and even plotlines. Properties percolate up to be able to hierarchically cover both lower- and higher-level structures. This is shown in short form between Figures 3 and 4—primary structures can be absorbed into the more important heads of phases. VNG predicts this to occur at larger levels as well, such that primary Peaks at the shot-level should also provide key information for the plot-level narrative. For example, just as the structures in Figure 3 can "percolate up" into Figure 4, this entire sequence plays a role within the broader narrative structure of the movie. This sequence is likely one among many embedded Peaks that constitute the Establisher of the overall story (i.e., it shows the Peak of the scene where the Jawas capture R2-D2, which thereby contributes to the "set up" of Luke acquiring the droids, and thereby facilitates his larger adventure).

This type of recursion is perhaps more apparent when it uses obvious structured patterns, and similar observations of multilevel recursive patterning have been made for Metz's patterns specifically (Fledelius 1978). For example, the final, climatic sequence of *Star Wars* shifts between scenes of the starfighters attempting to blow up the Death Star and the Rebellion fearing destruction at its hands. This multitracking does not occur at the shot level, but rather at the scene level. Thus, in addition to constructions manifesting as a surface structure to a sequence, VNG is able to capture observations about hierarchy at different levels.

Categorical roles

Film (and narrative) theory has long recognized that plotlines are organized into narrative "acts" that function in different ways. Such observations extend back to Aristotle's three-act

structure and Freytag's (1894) five-act "pyramid" for theatre, and, more recently, in a four-act structure theorized for the plots of mainstream Hollywood movies (Bordwell 2006, Thompson 1999). Similar structures have appeared in many models of narrative for both verbal, visual, and filmed narratives (for review, see Cutting Under Review, Cohn 2013b). Recent work has also shown that such plot-level structures are characterized by features of surface elements like shot pacing, cuts, and even motion and brightness in shots (Cutting Under Review). Such narrative roles are posited as psychologically reducing the processing costs of following a story (Cutting Under Review), and contribute towards visual *narratives* being distinguished from the dynamic updating of *semantic* information which could just as readily characterize the perception of everyday experience (Radvansky and Zacks 2014, Zacks 2014).

Insights of narrative roles throughout plotlines are maintained in VNG, and some preliminary work has suggested that VNG is effective for characterizing the shots of narrative data videos (Amini et al. 2015). Yet, as discussed above, the narrative categories in VNG also extend down to the level of scenes and shots, in addition to broader plotlines. This hierarchic nature allows VNG to capture additional observations about functional roles. Consider an ***establishing shot*** (Bordwell 2007, Bordwell and Thompson 1997, Cutting and Iricinschi 2015), which is a long shot showing an expansive view on both the environment and the people in it. It functions to situate elements of a scene prior to actions taking place, though experiments suggest that its absence does not necessarily impair the comprehension of a sequence (Kraft, Cantor, and Gottdiener 1991).

Establishing shots are clearly similar to Establishers in VNG, which also are rarely noticed when missing (Cohn 2014b). However, the notion of an "establishing shot" is more limiting that that of Establishers: they only account for shots that use a wide perspective on the whole scene. Yet, numerous images or shots may serve this same function, each showing only a glimpse of the overall "established" scene. For example, in *Star Wars* when Luke enters the Mos Eisley cantina, numerous shots show various aliens in acts of disrepute—all establishing the broader environment. Such a sequence is covered by Metz's patterns, but simply noting this as a relational syntagma misses the fact that these independent shots provide the same function as would a single establishing shot, only here distributed across multiple shots. VNG would label all of these as Establishers, but conjoined together using E-Conjunction. In addition, in VNG, such a Metzian surface pattern (E-Conjunction) could manifest in other conjoined narrative categories. We might therefore think of establishing shots as "prototypical single unit Establishers," though not all Establishers might be used as establishing shots.

Thus, in sum, because of its hierarchic, recursive properties, VNG allows us to characterize narrative roles at levels of the shot, scene, or plotline. Further, basic modifiers in VNG like conjunction characterize functional similarities across different shots/syntagma. Such features allow VNG to capture numerous features of film theories incorporated directly within the basic properties of the model.

VNG and Filmic Narrative Grammar

Finally, we can now ask: how does VNG change when describing moving images versus static images? There are many clear differences between static visual narratives, like those found in comics, and moving visual narratives, like those found in films. A basic prediction may be that, should differences arise between the structure of static and moving image sequences, they would be a direct result of the differences in the modalities themselves (Cohn 2013a). We now turn to exploring such potential differences.

Form and vocabulary

First, obvious differences emerge because static visual narratives are drawn while films are usually captured by a camera. Drawings are a trait of human biology—all children draw, whether it is using tools (pens, crayons) or just their fingers (in sand, paint, etc.), though tempered by cultural exposure to graphic patterns, just like language (Cohn 2012, Wilson 1988). Thus, drawn visual narratives manifest using a systematized graphic structure of patterned lines and shapes. Film does not use such producible patterns of an expressive modality (drawing), but rather is mediated by technology that is required for both its production (creating film) and reception (viewing film). In this way film does not require storage of graphic patterns, but rather bypasses it for a direct interface to general perception (except in animation, discussed below).

Despite this, it is worth remembering that most filmic sequences usually begin as drawn sequences: films are often preceded by storyboards which are static sequential images comparable to those found in comics. Thus, the *production* of meaningful image sequences in film usually uses static visual sequences as their basis, though their end-state emerges with perceptual motion within and between these static units, along with multimodal information of music, speech, etc.[4] Because of this, the narrative structure of film uses an inherently hybrid form. Film captures perceptual information through a camera; this alone would be comprehended through general perceptual principles and semantic understandings related to event knowledge. This information is then broken up into shots and edited together using a filmmaker's cognitive combinatorial narrative principles—here posited as Visual Narrative Grammar (often first produced in storyboards)—such that those units progress in a coherent sequence. The resulting "Filmic Narrative Grammar" is thus a hybrid between the natural VNG and its manifestation using units of general perception.

Unlike drawing, film's hybrid status is not natural. Static drawings come from the universal human ability to create meaning graphically. Instead, film captures perceptual information from an external source by filming it with a *specific* tool—a camera—and is thus mediated by technology developed only recently. FNG moves one step closer to VNG in *animation*, where the units are once again drawn. However, animation with motion may change VNG just as FNG does, by virtue of the affordances of moving images (discussed below).

One interesting aspect of this difference is that drawn narratives are not bound by contiguous motion across images (Zacks and Magliano 2011, Smith 2012, Berliner and Cohen 2011). Rather, drawn images often use discontinuous images where no "smooth motion" is carried from one image to the next. For example, a scene may show a person walking behind a tree and coming out the other side. Continuity editing would cut the scene when the person disappears behind the tree, so that their emergence from the other side looks fluid and continuous, but a discontinuous edit might stop the first shot before the person reaches the tree, only to cut to them coming out the other side.

In film, such discontinuity between shots alone should be jarring and less coherent because it deviates from our expectation of perceptual continuity (Zacks and Magliano 2011, Smith 2012, Berliner and Cohen 2011). Yet, perhaps more discontinuous image relations can persist in drawn narratives because they select only discrete moments out of an inferred continuous action and are not tied to perceptual motion—only *conceptual* motion. Or, perhaps

[4] Note that the "parallel architecture" approach in which VNG is embedded is also ideal for dealing with the multimodal characteristics of film. Though not focused on here, this model allows different structures to independently operate, yet to coalesce to create emergent structures and meaning not found in any one modality (Cohn 2016).

discontinuity between actions is actually acceptable in drawn form, but that film continuity gives a "bonus" for comprehension by further tying them together through motion? Alternatively, perhaps this is modality specific: Maybe discontinuity is acceptable only for the drawn form, but general perception in film leads to expectation of continuity? This would predict that discontinuous shots would be more acceptable in animation, and that lack of continuity would be less acceptable for static narratives using general perception, as in "photo comics" which are not drawn. While discontinuous shots likely do violate the perceptual continuity in film (as has been claimed), the relations to static, drawn narratives are worth exploring through empirical testing.

Insofar as film uses general percepts, it usually lacks the "vocabulary" of drawn visual narratives. For example, the visual language used in comics uses many conventionalized patterns like motion lines to imply the path of moving objects, bubbles and balloons to depict thoughts and speech, star shaped flashes to indicate impacts, and many other patterns, which are often culture specific (Cohn 2013a). In general, films do not use these conventions unless they are animated or trying to reference the style of comics (as in the film adaptation of the comic *Scott Pilgrim vs. the World*).

Another dimension of difference between filmed and drawn narratives, at least those in comics, is the ***layout*** of panels on a page. Filmed narratives typically involve no physical juxtaposition of shots, with images proceeding temporally in the single space of the screen. Granted, some films do play with the screen space to divide it up into component "panels" through split screens or more complicated framings, as in Ang Lee's *Hulk* (Bateman and Veloso 2013). However, "filmic panels" and page layouts in comics differ in important ways. First, the content in filmic panels usually occur simultaneously in real time, thereby implying narrative simultaneity, while those in static form may or may not be at simultaneous moments based on the particulars of the sequence. Second, there is no "reading order" for filmic panels on a screen independent of the relations between their content (in part due to their simultaneity and internal motion); ordering may be motivated by a variety of factors, including when those filmic panels appear, how they move around on the screen, and other aspects of internal content such as sound. In contrast, the ordered navigation of static images is integral to their comprehension, yet such ordering is provided by a rule system that is external to the content within the frames (Cohn 2013a, Cohn and Campbell 2015).

Narrative structure

Besides graphic structure, the clearest difference between drawn and filmic narratives is that film uses motion while drawings do not. This means that film is a continuous, inherently temporally progressing medium, while drawings form discrete units where the passage of time is ambiguous. For example, Figure 2b depicts three different characters in a sequence of conjoined Initials. In drawn form, nothing indicates the passage of time, and indeed all panels exist at the same narrative state (Initials), which further suggests simultaneity. Without explicit cues in the images, we cannot tell whether time changes. In contrast, if this sequence appeared in film, these shots would all inherently extend across a span of time, and thus would be more suggestive of time passing between them as well.

Additional consequences of motion may impact the units of the grammar. Because a camera can just be left recording, with no cuts, this temporality should be the equivalent of general perception of events. Such an ongoing recording of information alone has no *narrative* segmentation (discussed further below), though it may be comprised of perceived events, which do have their own segmental structure (Zacks et al. 2001) that interface with the narrative. When

that continuous perception is broken up into shots and edited together, the shots take on narrative categories linked through principles of FNG. However, because the units are now in motion, this temporality may "gloss over" features that in the static form would be individuated categories. Thus, the more cuts, and the less event information per cut, the more narrative units will be used in a film.

Consider shots 2, 4, 6, 11, and 12 in Figure 3. In all of these cases, I have chosen to show the shots by using multiple static frames. This is because the internal structure of these shots uses movement: in shot 2 the Jawas recede into shadows, in 4 they dart away in the shadows, in 6 a Jawa leaps up and shoots a blaster, and in 12 R2-D2 lurches back and falls over. In the tree structure in Figure 3, each shot has been notated for a narrative category wherein that movement occurs (e.g., shot 12 is a Release). Yet, if we were to consider Figure 3 as depicted—as static images and not as continuous shots—each image would have its own narrative category. For example, 12.1 would be an Initial of R2-D2's starting state before falling in the Peak of 12.2, which would then be contained within a Release *constituent*. As stated above, continuous shots contain no internal narrative structure, only event structures. Thus, the motion of film shots may "gloss over" what could otherwise be individuated units of the grammar in the drawn form, instead mapping to a higher level of structure.

In addition, not only do the elements within a film shot move (i.e. characters and objects move around), but so can the camera itself, with panning and zooming. For example, panning allows a camera to navigate through a space, while only showing a portion at a time, and/or retaining focus on certain moving elements in that space. Drawings can attempt to simulate panning with or without continuous motion, such as "divisional" panels that divide a singular background into different panels (Cohn 2014a). Figure 5 shows divisional panels that depict a moving bee. In graphic form, this illusion of motion will always retain some segmentation by virtue of the fact that the units do not inherently move.

Figure 5. Polymorphic Divisional panel of a bee flying

Another example is zooming. A person may be shown in full body, only to have the camera zoom into a close up of their eyeball in one continuous shot. In static form, a single panel of a full body followed by a close up panel on their eye would be a Refiner. However, in film, the continuity of the zoom makes the "head" category bleed into the "Refiner." Imagine if a shot zoomed from a fully body shot into a close up of an eye, then back to the full body. Would the zoom retain a "special" status as a Refiner (and what would that thereby do the previous state?), or would the continuous motion merely gloss over such structure as it does for the continuous

events? A similar issue occurs with Perspective Shifts, which merely alter the viewpoint on a scene from one image to another. The famous "bullet-time" scene from *The Matrix* captures this well: the camera sweeps in a circle around the frozen character of Neo, showing the whole scene from 360°. In graphic form, this would require several discrete Perspective Shift panels, but in film it would at least be possible to accomplish with a single shot.

Another example applies to the construction of an environment. In the second shot immediately following the end of the scene in Figure 3, the camera pans from R2-D2 lying on the ground to the standing Jawas, leaving R2-D2 out of the frame. As in the E-Conjunction in shots 7 and 8 in Figure 3, R2-D2 and the Jawas are never shown on screen at the same time. However, the continuity of this single film shot (and the background) likely "dampens" the inferential demand compared to that in unitized panels/shots. In film, this depicted contiguity between characters shows us that they belong to a single environment. With discrete images, this must be entirely inferred or must use a "divisional" panel as in Figure 5. Thus, for all aspects of the narrative grammar, from basic narrative roles to modifiers, the introduction of motion can mitigate or wash over what in static form might be clearly discrete units.

Thus far, I have claimed that the continuous motion of film leads FNG to "gloss over" structure that would otherwise be individuated in VNG when drawn. A strong view would thus posit that cuts between shots are necessary to individuate the component parts of the narrative grammar. Without such divisions a shot merely shows the progression of event structure, but not narrative. An alternative perspective though might say that cuts are *not* necessary to divide components, and event structures can map to narrative states within moving shots. This would be consistent with psychological theories of event cognition, which posit the same segmental structures within everyday events as those signaled by filmic cuts (Radvansky and Zacks 2014, Zacks 2014, Zacks and Magliano 2011), but which do not differentiate event knowledge from narrative structures (as in VNG).

Certainly, narrative structures *must* appear in some continuous percepts, or else no sense of narrative would emerge from live theatre other than hard scene breaks. However, this view would then need to explain the difference between general perception of everyday events and filmed narratives—which are clearly not the same phenomenon. Conversely, not all cuts must constitute cues for *narrative* structure. For example, the filming of live events like sports, concerts, political debates, and others also use cuts to change camera angles, show different people, etc. Yet, each cut within such filming does not signal a narrative break—indeed, such events are usually not narratives at all! Thus, cuts alone are not signals for narrative categories, since shots may not even have narrative categories.

There may also be a middle ground in the case of narrative films. For example, perhaps narrative structures—which are different from just event knowledge (Cohn 2015b)—appear in continuous shots only when they depart from general percepts and/or use cues to signal the changing between narrative units. Thus, there would be no component narrative categories for simple continuous events (i.e., in Figure 3, shot 12 would not divide into an Initial-Peak constituent), but such structure would arise for continuous zooms for Refiners or panning for Perspective Shifts (like the *Matrix*'s bullet-time). These are conventions of films that do not occur in everyday perception, and thus would be "marked" as necessitating narrative structures. Films may also use transitions to tell us that they are continuously shifting from one narrative category to another. Imagine an expansive establishing shot of characters in a scene (an Establisher), which then zooms in one motion to the characters involved in actions within that

scene (Initial, etc.). In this case, the zoom may functionally tell us about the change from one category to another, without a cut. Disentangling these issues remains an empirical question.

Summary

Overall, this section has outlined several ways in which the structure of static, drawn narratives differs from that of moving, filmic narratives. Such departures are hypothesized as primarily emerging from the differences in basic properties of the media themselves (static vs. moving). Some of these differences are summarized in Table 1.

Table 1. Gross differences in dimensions between prototypical cases of drawn and filmed narratives

Static, drawn narratives	**Moving, filmic narratives**
Production is a biologically based human ability (drawing)	Production is technologically mediated (non-natural)
Uses patterned graphic schema for both iconic and symbolic elements (i.e., stored lines and shapes in a visual vocabulary)	Uses general perception (not a patterned visual vocabulary, with the exception of animation)
Static content in images	Moving content in film
Static depictions in images	Moving camera in film (panning, zooming)
Ambiguous temporality between units unless otherwise depicted	Pervasive sense of temporality between units because of ongoing temporality of motion
Spatial juxtaposition of units (in page layout) requiring non-content based navigational rules	Temporal juxtaposition of units unfurling on a screen. When spatially juxtaposed frames appear on a screen, they involve no independent navigational rules.

Conclusion

Altogether, I have introduced a theory of narrative grammar that makes numerous insights made by previous theories of visual narratives, such as narrative categories, hierarchic structures, and patterned constructions. In its application to static visual narratives, this architecture has growing experimental support for its psychological validity. Insofar as the brain relies on domain general mechanisms for narrative comprehension, such an approach has been posited as capable of describing film comprehension as well, and work has begun to examine such applicability (e.g., Amini et al. 2015). However, insofar as static and moving narratives create different affordances, this narrative grammar alters in structure for the hybrid nature of film's moving images compared to the static, more natural drawn modality. This theory thus allows for specific predictions about—and the ability to test—both the comprehension of FNG and the differences between VNG and FNG through empirical investigation.

References

Amini, Fereshteh, Nathalie Henry Riche, Bongshin Lee, Christophe Hurter, and Pourang Irani. 2015. "Understanding Data Videos: Looking at Narrative Visualization through the Cinematography Lens." In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 1459-1468. ACM.

Bateman, John A. 2007. "Towards a grande paradigmatique of film: Christian Metz reloaded." *Semiotica* 2007 (167):13-64.

Bateman, John A., and Karl-Heinrich Schmidt. 2012. *Multimodal film analysis: How films mean*. New York: Routledge.

Bateman, John A., and Francisco O. D. Veloso. 2013. "The semiotic resources of comics in movie adaptation: Ang Lee's Hulk (2003) as a case study." *Studies in Comics* 4 (1):135-157. doi: 10.1386/stic.4.1.135_1.

Bateman, John A., and Janina Wildfeuer. 2014. "A multimodal discourse theory of visual narrative." *Journal of Pragmatics* 74:180-208. doi: 10.1016/j.pragma.2014.10.001.

Berliner, Todd, and Dale J Cohen. 2011. "The illusion of continuity: Active perception and the classical editing system." *Journal of Film and Video* 63 (1):44-63.

Bordwell, David. 2006. *The way Hollywood tells it: Story and style in modern movies*. Berkeley, CA: University of California Press.

Bordwell, David. 2007. *Poetics of Cinema*. New York, NY: Routledge.

Bordwell, David, and Kristin Thompson. 1997. *Film Art: An Introduction*. 5th Edition ed. New York, NY: McGraw-Hill.

Branigan, Edward. 1992. *Narrative Comprehension and Film*. London, UK: Routledge.

Buckland, Warren. 2000. *The Cognitive Semiotics of Film*. Cambridge: Cambridge University Press.

Byram, M.L., and Chris Garforth. 1980. "Research and testing non-formal education materials: a multi-media extension project in Botswana." *Educational Broadcasting International* 13 (4):190-194.

Carroll, John M. 1980. *Toward a Structural Psychology of Cinema*. The Hague: Mouton

Carroll, John M., and Thomas G. Bever. 1976. "Segmentation in cinema perception." *Science* 191 (4231):1053-1055.

Chateau, Dominique. 1987. *La cinéma comme langage*. Brussels: AISS - Publications de la Sorbonne.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Cohn, Neil. 2010. "The limits of time and transitions: Challenges to theories of sequential image comprehension." *Studies in Comics* 1 (1):127-147.

Cohn, Neil. 2012. "Explaining "I can't draw": Parallels between the structure and development of language and drawing." *Human Development* 55 (4):167-192. doi: 10.1159/000341842.

Cohn, Neil. 2013a. *The visual language of comics: Introduction to the structure and cognition of sequential images*. London, UK: Bloomsbury.

Cohn, Neil. 2013b. "Visual narrative structure." *Cognitive Science* 37 (3):413-452. doi: 10.1111/cogs.12016.

Cohn, Neil. 2014a. "The architecture of visual narrative comprehension: The interaction of narrative structure and page layout in understanding comics." *Frontiers in Psychology* 5:1-9. doi: 10.3389/fpsyg.2014.00680.

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

Cohn, Neil. 2015a. "How to analyze visual narratives: A tutorial in Visual Narrative Grammar." http://www.visuallanguagelab.com/P/VNG_Tutorial.pdf.

Cohn, Neil. 2015b. "Narrative conjunction's junction function: The interface of narrative grammar and semantics in sequential images." *Journal of Pragmatics* 88:105-132. doi: 10.1016/j.pragma.2015.09.001.

Cohn, Neil. 2016. "A multimodal parallel architecture: A cognitive framework for multimodal interactions." *Cognition* 146:304-323. doi: 10.1016/j.cognition.2015.10.007.

Cohn, Neil, and Patrick Bender. Under Review. "Drawing the line between constituent structure and coherence relations in visual narratives." *Journal of Experimental Psychology: Learning, Memory, & Cognition*.

Cohn, Neil, and Hannah Campbell. 2015. "Navigating comics II: Constraints on the reading order of page layouts." *Applied Cognitive Psychology* 29:193-199. doi: 10.1002/acp.3086.

Cohn, Neil, Ray Jackendoff, Phillip J. Holcomb, and Gina R. Kuperberg. 2014. "The grammar of visual narrative: Neural evidence for constituent structure in sequential image comprehension." *Neuropsychologia* 64:63-70. doi: 10.1016/j.neuropsychologia.2014.09.018.

Cohn, Neil, and Marta Kutas. 2015. "Getting a cue before getting a clue: Event-related potentials to inference in visual narrative comprehension." *Neuropsychologia* 77:267-278. doi: 10.1016/j.neuropsychologia.2015.08.026.

Cohn, Neil, and Marta Kutas. In prep. "What's your function, narrative conjunction? Structure, meaning, and experience in the neurocognition of visual narratives."

Cohn, Neil, and Martin Paczynski. 2013. "Prediction, events, and the advantage of Agents: The processing of semantic roles in visual narrative." *Cognitive Psychology* 67 (3):73-97. doi: 10.1016/j.cogpsych.2013.07.002.

Cohn, Neil, Martin Paczynski, Ray Jackendoff, Phillip J. Holcomb, and Gina R. Kuperberg. 2012. "(Pea)nuts and bolts of visual narrative: Structure and meaning in sequential image comprehension." *Cognitive Psychology* 65 (1):1-38. doi: 10.1016/j.cogpsych.2012.01.003.

Cohn, Neil, and Eva Wittenberg. 2015. "Action starring narratives and events: Structure and inference in visual narrative comprehension." *Journal of Cognitive Psychology* 27 (7):812-828. doi: 10.1080/20445911.2015.1051535.

Colin, Michel. 1995. "The Grande Syntagmatique revisited." In *The Film Spectator: From Sign to Mind*, edited by Warren Buckland, 45-86. Amsterdam: Amsterdam University Press.

Culicover, Peter W., and Ray Jackendoff. 2005. *Simpler Syntax*. Oxford: Oxford University Press.

Cutting, James E. Under Review. "Narrative theory and the dynamics of popular movies." *Psychonomic Bulletin & Review*.

Cutting, James E., and Catalina Iricinschi. 2015. "Re-Presentations of Space in Hollywood Movies: An Event-Indexing Analysis." *Cognitive Science* 39 (2):434-456. doi: 10.1111/cogs.12151.

Eisenstein, Sergei. 1942. *Film Sense*. Translated by J. Leyda. New York, NY: Harcourt, Brace World. Original edition, 1923.

Fledelius, Karsten. 1978. "Syntagmatic Film Analysis–with Special Reference to Historical Research." *Untersuchungen zur Syntax des Films I*:31-68.

Freytag, Gustav. 1894. *Technique of the Drama*. Chicago: S.C. Griggs & Company.

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

Gernsbacher, Morton Ann. 1985. "Surface information loss in comprehension." *Cognitive Psychology* 17:324-363.

Goldberg, Adele. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago, IL: University of Chicago Press.

Hagmann, Carl Erick, and Neil Cohn. 2016. "The pieces fit: Constituent structure and global coherence of visual narrative in RSVP." *Acta Psychologica* 164:157-164. doi: 10.1016/j.actapsy.2016.01.011.

Ildirar, Sermin, and Stephan Schwan. 2015. "First-time viewers' comprehension of films: Bridging shot transitions." *British Journal of Psychology* 106 (1):133-151. doi: 10.1111/bjop.12069.

Jackendoff, Ray. 1990. *Semantic Structures*. Cambridge, MA: MIT Press.

Jackendoff, Ray. 2002. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

Kraft, Robert N., Phillip Cantor, and Charles Gottdiener. 1991. "The coherence of visual narratives." *Communication Research* 18 (5):601-616.

Kunzle, David. 1973. *The History of the Comic Strip*. Vol. 1. Berkeley: University of California Press.

Levin, Daniel T., and Daniel J. Simons. 2000. "Perceiving Stability in a Changing World: Combining Shots and Intergrating Views in Motion Pictures and the Real World." *Media Psychology* 2 (4):357-380.

Magliano, Joseph P., Adam M. Larson, Karyn Higgs, and Lester C. Loschky. 2015. "The relative roles of visuospatial and linguistic working memory systems in generating inferences during visual narrative comprehension." *Memory & Cognition*:1-13. doi: 10.3758/s13421-015-0558-7.

Magliano, Joseph P., and Jeffrey M. Zacks. 2011. "The Impact of Continuity Editing in Narrative Film on Event Segmentation." *Cognitive Science* 35 (8):1489-1517. doi: 10.1111/j.1551-6709.2011.01202.x.

Mandler, Jean M., and Nancy S. Johnson. 1977. "Remembrance of things parsed: Story structure and recall." *Cognitive Psychology* 9:111-151.

Marslen-Wilson, William D. 1987. "Functional parallelism in spoken word-recognition." *Cognition* 25 (1-2):71-102.

Martin, James R. 1983. "Conjunction: The logic of English text." In *Micro and Macro Connexity of Discourse (= Papers in Textlinguistics 45)*, edited by J.S. Petöfi and E. Sözer, 1-72. Hamburg: Helmut Buske Verlag.

McCloud, Scott. 1993. *Understanding Comics: The Invisible Art*. New York, NY: Harper Collins.

Metz, Christian. 1974. *Film Language: A Semiotics of the Cinema*. Translated by Michael Taylor. New York: Oxford University Press.

Miller, George A., and Stephen Isard. 1963. "Some perceptual consequences of linguistic rules." *Journal of Verbal Learning and Verbal Behavior* 2 (3):217-228. doi: 10.1016/s0022-5371(63)80087-0.

Osterhout, Lee, and Janet L. Nicol. 1999. "On the Distinctiveness, Independence, and Time Course of the Brain Responses to Syntactic and Semantic Anomalies." *Language and Cognitive Processes* 14 (3):283-317.

Radvansky, Gabriel A., and Jeffrey Zacks. 2014. *Event Cognition*. Oxford, UK: Oxford University Press.

Schwan, Stephan, and Sermin Ildirar. 2010. "Watching Film for the First Time: How Adult Viewers Interpret Perceptual Discontinuities in Film." *Psychological Science* 21 (7):970-976. doi: 10.1177/0956797610372632.

Smith, Tim J. 2012. "The attentional theory of cinematic continuity." *Projections* 6 (1):1-27. doi: http://dx.doi.org/10.3167/proj.2012.060102.

Thompson, Kristin. 1999. *Storytelling in the new Hollywood: understanding classical narrative technique*. Cambridge, MA: Harvard University Press.

van Leeuwen, Theo. 1991. "Conjunctive structure in documentary film and television." *Continuum* 5 (1):76-114. doi: 10.1080/10304319109388216.

Wildfeuer, Janina. 2013. *Film Discourse Interpretation: Towards a New Paradigm for Multimodal Film Analysis*. New York: Routledge.

Wilson, Brent. 1988. "The Artistic Tower of Babel: Inextricable Links Between Culture and Graphic Development." In *Discerning Art: Concepts and Issues*, edited by George W. Hardiman and Theodore Zernich, 488-506. Champaign, IL: Stipes Publishing Company.

Wilson, Brent. 2016. "What happened and what happened next: Kids' visual narratives across cultures." In *The Visual Narrative Reader*, edited by Neil Cohn, 185-227. London: Bloomsbury.

Wilson, Brent, and Marjorie Wilson. 1987. "Pictorial Composition and Narrative Structure: Themes and Creation of Meaning in the Drawings of Egyptian and Japanese Children." *Visual Arts Research* 13 (2):10-21.

Zacks, Jeffrey M. 2014. *Flicker: Your Brain on Movies*. Oxford, UK: Oxford University Press.

Zacks, Jeffrey M., Todd S. Braver, Margaret A. Sheridan, David I. Donaldson, Abraham Z. Snyder, John M. Ollinger, Randy L. Buckner, and Marcus E. Raichle. 2001. "Human brain activity time-locked to perceptual event boundaries." *Nature Neuroscience* 4 (6):651-655.

Zacks, Jeffrey M., and Joseph P. Magliano. 2011. "Film, narrative, and cognitive neuroscience." In *Art and the Senses*, edited by D.P. Melcher and F. Bacci, 435-454. New York: Oxford University Press.

Zacks, Jeffrey M., Nicole K. Speer, Khena M. Swallow, and Corey J. Maley. 2010. "The brain's cutting-room floor: Segmentation of narrative cinema." *Frontiers in Human Neuroscience* 4:1-15.

Zwaan, Rolf A., and Gabriel A. Radvansky. 1998. "Situation models in language comprehension and memory." *Psychological Bulletin* 123 (2):162-185.