Studies in Comics
Volume 1 Number 1
 doi: 10.1386/stic.1.1.127/1

NEIL COHN
Tufts University

The limits of time and transitions: challenges to theories of sequential image comprehension

Keywords
visual language
panels
time
temporal map

Abstract

The juxtaposition of two images often produces the illusory sense of time passing, as found in the visual language used in modern comic books, which creates the sense that this linear sequence presents a succession of moments or temporal units. Author and theorist Scott McCloud took this view to an extreme, proposing that sequential images are guided by a notion that 'time = space' (McCloud 2000), because this temporal passage occurs on a spatial surface. To McCloud, this 'temporal mapping' results in a movement of time with a movement of space. This sense of temporality, then, is the 'essence' of comics, which is manifested in McCloud's taxonomy of transitions of panel-to-panel relationships (McCloud 1993). While less specific, this same type of 'essence' of connection can be reflected in Groensteen's types of 'arthrology' across a linear sequence or disparate panels in a broader text (Groensteen 1999).

However, numerous problems arise with McCloud and Groensteen's approaches to graphic narrative. This article will explore how the linearity of reading panels and the iconicity of images create various false assumptions about the conveyance of meaning across sequential images' depictions of space and time. With numerous examples, it will argue that any linear panel-to-panel analysis (such as McCloud's (1993) panel transitions) or loosely defined principles of connection (such as Groensteen's (1999) 'arthrology') between sequential images are inadequate to account for their understanding. The conclusion is that sequential image comprehension must be thought of as the union of conceptual information that is grouped via unconscious hierarchic structures in the mind. As such, the study of the comprehension of the visual language used in comics must be placed in the cognitive sciences.

Introduction

Any discussion of structure and meaning in the visual language used in comics must contend with the central issue of the comprehension of sequential images. Popular research on this issue has taken a viewpoint similar to the phenomenological experience of reading – since each panel is read (or drawn) in succession, they must be understood in succession. This perspective has been formalized in theories of 'transitions' between panels (e.g., McCloud 1993), while a strong version proposes that each of these images represents some successive temporality, and spatial mapping to a progression in time (e.g., McCloud 1995; 2000). Finally, a third variation on this theme takes the position that individual panels do not just connect with their linear neighbours, but are multiple to other panels throughout a physical document (i.e., a strip, page, book, etc.) (e.g., Groensteen 1999). This article will argue that all of these orientations are ultimately unable to describe how meaning is created in sequential images, and that any theory aiming at such a goal needs to take into account complex groupings of panels motivated by principles in the human mind.

Meaning?

In exploring meaning in sequential images, we first ask 'where does meaning come from?' Since comic author/theorist Scott McCloud (1993) and scholar Thierry Groensteen (1999) have become the most popular voices on this topic, and both compare the structure of comics to language, it is perhaps worth comparing their views to the linguistic treatment of meaning. Groensteen largely treats the meaning of sequential images as a product of culture – a 'system' out in the world that readers access, lining up with his semiological roots in European comics scholarship (e.g., Gubern 1972, Hünig 1974, Koch 1971, Krafft 1978). However, these 'structuralist' views pointing to the origin of meaning as a product of culture have been largely dismissed by mainstream linguistics for the past fifty years (Harris 1993).

Rather, since the 'cognitive revolution' in psychology and linguistics in the middle of the twentieth century, meaning has been recognized as stemming wholly from the mind (Jackendoff 1983; 1987). Following this view, any discussion of the structure of sequential images' meaning is really a discussion about the mechanisms in the mind/brain that guide comprehension of such outwardly manifested phenomena.[1] McCloud quite admirably taps into this insight, with much of his theories directly invoking a 'reader's involvement' in comprehension. This terminology leaves some (possibly appealing) ambiguity to the issue though: it is unclear when he talks about 'reader participation' or 'involvement' whether he means to say that these mental processes are conscious or unconscious.

This article will argue that the processes guiding sequential image comprehension remain inaccessible to conscious awareness. To this point, the sense in which 'meaning' is explored has no concern with any conscious sense of 'artistic interpretation'. Instead it is about the basic comprehension of sequential images – comparable to the creation of meaning by words in sentences. While we are consciously aware of (sometimes various) meanings of sentences, the unconscious processes that motivate these understandings remain inaccessible. The comprehension of sequential images is taken to work in the same way.

Furthermore, this ability appears not to be transparent and universally accessible, and requires a degree of expertise. Studies have shown that while all Japanese 6 year olds (who have high exposure to manga) can draw sequential narratives, less than half of 12 year olds in other countries (who have low exposure to comics) could create a coherent pairing of juxtaposed panels (Wilson 1999, Wilson and Wilson 1987). Other studies have indicated that comprehension ability for sequential images correlates with age and expertise (Nakazawa 2005, Pallenik 1986). This 'expertise' must be affecting something in the mind – not just conscious awareness, and it is those mechanisms that this piece will explore.

Sequential images

The most prevalent belief about the understanding of sequential images – or 'panels' – holds that comprehension progresses in a linear fashion, similar to the reading process itself. Like approaches to syntax of spoken languages prior to the 1950s, this linear approach has been codified in taxonomies of 'transitions', which specify the nature of the shift from one panel to another. While other breakdowns have been offered (Christiansen 2000), McCloud's analysis (McCloud 1993: 70–72) gained popularity and influence beyond most others, including expansions building from McCloud's own work (such as Cohn 2003, Dean 2000, Saraceni 2000; 2003, Stainbrook 2003). McCloud identified six transitions to characterize the relations of one panel to another:

1. Moment-to-moment – between small increments of time
2. Action-to-action – between full ranges of actions

1. Languages 'out in the world' in this cognitive view merely become an idealized average across mutually intelligible patterns in individuals' heads. In other words, a language out in the world (i.e., 'English' or 'Swahili') refers to an 'idealized average' of the patterns that are shared in the minds/brains of a particular population of people (Chomsky 1986). Such a view is easily transferable to sequential images, as found in comics (Cohn 2003; 2005).

3. Subject-to-subject – between characters or objects in a scene
4. Aspect-to-aspect – between aspects of a scene or an environment
5. Scene-to-scene – between different scenes
6. Non-sequitur – have no apparent meaningful relation

To enable these transitions, McCloud invokes the process of 'closure', where the mind 'fills in the gap' between images. McCloud states

> Comics panels fracture both time and space, offering a jagged staccato rhythm of unconnected moments. But closure allows us to connect these moments and mentally construct a continuous unified reality.
>
> (McCloud 1993: 67)

With closure, McCloud seeks out a cognitive mechanism to explain his theory of transitions. He astutely senses the insufficiency of simply saying that a transition is there – it must be facilitated by something in a reader's mind. For McCloud, this heralds an interaction between the meaning 'on the page' and the invisible meaning created in the mind between panels.

McCloud later refined his approach to what he called the 'essence of comics' – the 'temporal map'. He argues that by physically moving from one panel moment to another, space *equals* time (McCloud 1995; 2000). With this underlying equation, factors such as panel sizes and the distance between panels provide modifications on the understanding of time in visual form: the longer the panel or gap between panels, the longer the progression of time (McCloud 1993). Temporal mapping is a more radical version of transitions, since instead of panels just making meaningful connections of various types, there is a *presumed* temporality overarching across all meaning. It also washes over previous distinctions in transitions that differentiated between temporally progressive and temporally ambiguous panel relationships (as subcategorized in Cohn 2003). For example, subject-to-subject and aspect-to-aspect transitions do not inherently make a temporal distinction – they are largely about shifts in space within an environment – which McCloud himself attributes to their narrative advantage. However, with temporal mapping in place, all transitions show shifts in time, no matter what additional distinctions in content are highlighted.

Panel transitions and temporal mapping as theoretical concepts can make a lot of sense. Experientially, when reading a comic page, we engage the sequence one panel at a time. And, after all, we feel as though we live our lives through a series of moments, one after another. Why would juxtaposed images not mirror this experience of time?

Temporality and sequential images

An initial problem with the temporal mapping thesis is that the properties of space and time being compared exist on two entirely separate planes of analysis. While the space McCloud refers to is based on physical distance, the sense of time he refers to is entirely a mental construct garnered from the contents of the panels. This mapping of physical reading to time is confirmable because of his instruction for the reader to run their finger along the page to get a sense of space (McCloud 2000: 206), and further revealed as he observes,

> Wherever your eyes are focused, that's now. But at the same time your eyes take in the surrounding landscape of past and future! Like a storm front, the eye moves over the comics page, pushing the warm, high-pressure future ahead of it, leaving the cool low-pressure past in its wake.
>
> (McCloud 1993: 104)

As he describes the act of reading (eyes moving along a page), he equates it to the comprehension that there is no future or past tense within the images, only the sense of those states in surrounding panels. However, this meaning is *fictive* time, not 'real' experiential time. The time it takes to read something and the mental abstraction of time within the fictitious narrative are not comparable, and exist on totally different levels of analysis and experience. This is why fictive time is unaffected by different arrangements of the same panels in varying layouts, though layout might affect the physical *rhythmic* pace in which those panels are read (though this has not been tested empirically either).

Though ill formed, the idea of temporal mapping is merely symptomatic of a much more pernicious issue: the belief that *panels* equal *moments*. McCloud himself acknowledges the duplicity of believing that a single panel equals a single moment in time, citing the problems created when integrating text into an image through the use of speech balloons (McCloud 1993: 96–97). Because text relates to speech – which must be experienced temporally – McCloud notes that balloons spoken by two separate people in the same panel must represent different instances in fictive time, insinuating its progression. Thus, a panel can contain the *duration* of time beyond just a single moment.

A similar phenomenon occurs when an entire event is represented in a single panel. These 'polymorphic' panels (Cohn 2007) – also known as 'stroboscopic' (Cutting 2002) – show a single entity repeated in multiple positions of an action while remaining in a single encapsulated frame (as in the final panel of figure 3). Again, matching McCloud's observation, these panels seemingly represent the duration of time, rather than a single instance where the entity would seem to be in multiple positions at the same moment.

However, there are distinct differences between McCloud's duration panels and polymorphic panels. McCloud's examples rely on the introduction of written language, which through association

to *sound* must take up time, while the sense of interval in polymorphic panels comes directly from the content of the images. This difference is significant, particularly because of the entailment that McCloud makes from his example. The reading of balloons generally mimics the reading pattern of the whole composition, roughly reflecting the path of writing systems. In the case of English, a balloon that is further left in a panel is read first – thereby insinuating that it comes *before* those balloons following it to the right in fictive time (generally speaking, reading order mimics text to follow a preference hierarchy: higher > left > right > lower). In a polymorphic panel, this compositional restriction does not affect the representations of entities at different states.

Why is this difference important? McCloud's examples retain a sense of fictive temporal movement that matches the physical order of reading, thereby allowing him to uphold his equivalence of time and space. Under this interpretation, spatial shift still acceptably indicates temporal shift. Indeed, McCloud's temporal mapping hypothesis can be clarified further, making the equation not 'physical space = fictive time', but rather 'physical space = physical *reading motion* = fictive time'. McCloud must then reconcile these elements, stating, 'As readers, we're left with only a vague sense that as our eyes are moving through space, they're also moving through time – we just don't know by how much!' (McCloud 1993: 100).

Polymorphic panels contrast the temporal mapping equation because they allow for duration that is not attached to any sense of spatial progression. They purely show the passage of fictive time – regardless of their internal composition or how one might need to engage them physically in the act of reading. Figure 1 shows the final three panels of a strip by Kazu Kibiushi (2005), where the final polymorphic panel disallows strict temporal mapping, since it does not rely on left-to-right reading to understand the path of motion of the dog chasing his tail.

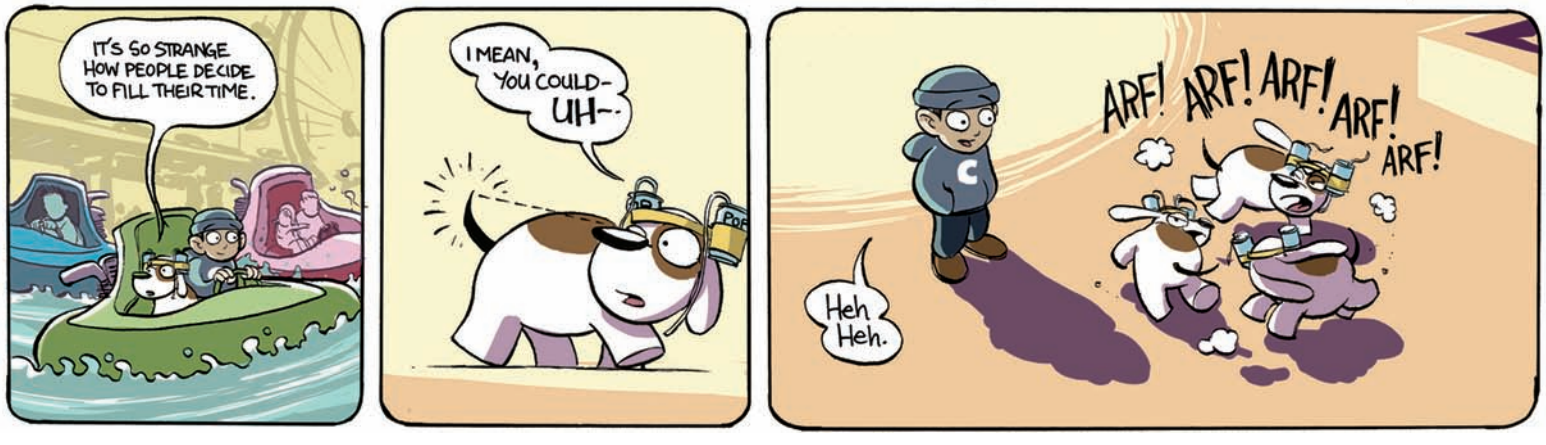

Figure 1: Polymorphic panel in Copper *by Kazu Kibiushi (2005).*

In fact, because of the dog's circular motion, the determination of a starting and ending point of motion becomes ambiguous, imbuing it with a recursive quality. Technically, this depiction has no start and has no end, and its duration of 'time' *never* stops, regardless of how the eye interacts with it as a panel on a page. Truly, this image does not depict a dog engaged in a temporally bound event that a reader's mind decides upon a start and ending for at all. Rather, it conveys the durative *concept* of a dog running in circles – no matter what temporal boundaries might be involved for that to take place.

A similar phenomenon occurs with the 'smoke-veiled fight', which conveys the concept of a fight without actually showing it, as in figure 2.

Figure 2: Smoke veiled fight.

The smoke-veiled fight is interesting in terms of temporality because we know that the event of fighting depicted has to be some duration of time. The representation does not allow it to be interpreted as a single instance of a punch or kick – it has to mean a series of combative actions. Nevertheless, the physical drawing never shows multiple actions or even any actions at all. Because of this, finding the start or end points of this duration is impossible, since the actions are not depicted. This meaning is entirely conceptual.

Finally, temporal mapping faces the problem that it implies that all spatial relations must then have temporal consequences. If this were true, how would it be quantified? If larger gaps between panels mean larger spaces of time, should all panels with large gaps in time be expected to have large physical gaps between them? Does more 'time' occur when needing to move diagonally between

panels in separate layout rows than those within a row? Is there thus even more time between panels when a page needs to be turned?

McCloud himself would acknowledge that such a rigid equation of physical time and fictive time is ridiculous. However, where is the line drawn? At what point does temporal mapping become active or not? How valid can temporal mapping be as a theory of comprehension if the boundaries of its application cannot be quantified? Does it only apply to explicitly temporal transitions like action-to-action or moment-to-moment panels, and if so, what need is there for an additional notion of temporal mapping? Such concerns necessarily turn us away from carrying a presumption that the passage of time is an overarching constraint on all sequential image meaning.

From time to concepts

All of these issues are related to a deeper problem: the notion that fictive time passes in or between images. Most likely, the pervasive belief in 'time' stems from a lack of acknowledgement that graphic images represent conceptual information. By and large, the images in the visual language of comics follow an *iconic* form of semiotic reference, because they *resemble* their meaning (Peirce 1931). Because we experience 'reality' on the same terms that we engage iconic images, we forget that they are indeed *representations* that come from – and must be processed by – a human mind: an observation harder to ignore with symbolic phenomena like spoken words.

Panels as units do not stand for moments or durations in fictive time, but direct attention to depictions of 'event states' (Cohn 2007) *from which* a sense of 'time' is derived. Images are just significations made meaningful through cognitively based concepts, while 'time' is a mental extraction from the causation/change between them. Indeed, nothing about two images next to each other demands that each represents a moment in 'time'. The entire sense of 'time' is pulled from the content of what the panels have in them. In other words, because two panels might depict states of an event – and because our knowledge of events is that they occur in the context of our perceived linear notion of time – we assume that 'time passes' between the two panels. However, there is no 'time' there, nor can any be assumed to be 'filled into the gaps' in any real semantic sense, unless information in the representation is presented to us. The binding assumption that each panel represents a moment or duration in time is merely an illusion, cast by the (unconscious) understanding of events and their parts.

Neither of the 'durations of time' discussed above need be analysed in temporal terms. The enclosure of speech balloons into panels does not mean that the time of 'real-life' speech content passes within the bounds of the image, but only that the *conceptual* content of the text is unified with that of the images. Meanwhile, polymorphic panels merely show the event states of a full action enclosed into a single bounded attention unit. As demonstrated with the polymorphic and

smoke-veiled fight examples, this does not have to be a sense of 'time' that ends or starts in a certain place, but can simply depict the conception of a durative action.

While it might seem like splitting hairs, the distinction between representation of time and event states remains important because it distinguishes which leads the dance in cognitive understanding of panels in sequence. While the theory of panel transitions does not stand in opposition to such conceptualism, the acknowledgment of this difference can lead to multiple advances beyond the linear analyses. First off, it shifts the focus from the *unseen* to the *seen*. Because McCloud's approach equates space and time, it assumes that only slices of an action are shown, the rest of which falls mysteriously between the panels. This 'gutter' *between* the panels thereby becomes the primary processing point for mental unification. McCloud states that

> The gutter plays host to much of the magic and mystery that are at the very heart of comics … in the limbo of the gutter, human imagination takes two separate images and transforms them into a single idea.
>
> (McCloud 1993: 66)

The 'mental' properties provided by linear transitions are, in part, just rhetorical hand waving. If closure occurs 'in the gaps between panels' then how does it work if a reader cannot make such a connection until the second panel is reached? That is, the gap cannot be filled unless it has already been passed over, making closure an additive inference that occurs *at* panels, not between them. Nowhere is this duplicity more apparent than in figure 3 and the text that accompanies it, which serves as McCloud's primary example for the process of closure between panels.

Figure 3: McCloud's primary example for closure.

> I may have drawn an axe being raised … but I'm not the one who let it drop or decided how hard the blow, or who screamed, or why. That, dear reader, was your special crime, each of you committing it in your own style. All of you participated in the murder.
>
> (McCloud 1993: 68)

Here, McCloud pulls a deft trick, since the ambiguous 'reader-created outcome' of the event cannot be attributed to the gutter, but to the *indexical quality* of the second panel's scream. It only implies that an event occurs through the indexical knowledge that speech functionally connects to a speaker, without demonstrating the assumed action (axe chopping), thereby making the outcome ambiguous. Thus, McCloud actually *does* control the depiction of the crime by *not showing it*. Really though, the gutter does not provide any meaning – the content of the panels and their union does. In this case, the conceptual basis of the images becomes even more salient, since the second panel does not even show the action it references, thereby heightening its inferential processes.

By accepting the conceptual basis of representations, the important focus for processing becomes the *content* of the panels. Indeed, individual panels must *also* derive their meanings from the mind of the reader, and, in some cases, that meaning relies on other panels in the sequence. Like in the metonymic example above, this allows for relational aspects of panels to be explored without the presumption of time restrictions, which begets discoveries that further invalidate any linear approach to understanding.

Beyond juxtaposition

Without the assumption that a linear flow of time dictates understanding, the very idea of linearity itself can be questioned. Take for example the sequence in figure 4. The final two panels are assumed to take place within the house depicted in the first panel. No indicators within those panels give us this information – it comes purely from the juxtaposition of the house as a locative in the preceding panel. There is no reason that this first panel should be considered as a 'moment' in time at all, since it functions entirely as a semantic locative. Moreover, that locative must hold scope over all of the panels it affects, not simply the panel immediately following it. That is, the second two panels must be grouped in a way to allow the first panel to apply to both of them.

Tapping into this idea of multiple connections, Barber (2002) proposed an intersection of transitions with page layouts. For him, panel transitions could extend across multiple panels and pages are read as a whole. However, there are problems with tying layout to meaning. First, so long as they are

Figure 4: Locative panel with no temporal relationship.

read in the same order, panels can be rearranged in different designs without changing meaning. The sequence of three panels in figure 4 could be read horizontally, vertically, diagonally, etc. However, changing those physical orders would make little alteration to meaning. Indeed, this is done frequently in newspaper comic strips, where editors might alter the layouts of strips to better fit their desired page layouts. The result then is that navigation of layout (i.e., deciding which panel to read next, usually glossed as 'left-to-right and down') is a separate system to the comprehension of sequential images. These components are no doubt interfaced, since layout may have various ways of influencing meaning (and possibly vice-versa), but they are not the same system.

Groensteen (1999) also accepts that panels interact not only with linear relations but also with relations to all other panels on a page. His principle of 'arthrology' extends this to include the connections between physical compositions of panels in a page or across pages. Arthrology is a broad concept, which on the one hand refers to aspects of *physical* composition – like 'visual rhyming' or a thematic leitmotif, as well as to his principle of 'braiding', which refers to 'threads' of meaning. Groensteen has stated that not only should one panel connect to another in a linear way, but that 'every panel exists, potentially if not actually, in relation with each of the others' (Groensteen 2007: 146).

However, if braiding serves as a model of comprehension, such unrestrained transitions (semantic relations between individual panels) between every possible panel in a document would overload the working memory of the human mind. To push this to an extreme: for an average book that has six panels per page for 24 pages, the 144 panels would potentially warrant connections between any two

panels, calculable as 144!/(2!•142!). This would yield 10,296 possible transitions for every combination possible, and the mind would have to retain each of these in memory additively with each successive panel read. Though not all relations may need to establish a connection, all transitions would be necessary to at least confirm or deny the need for an explicit transition. As Cohn (2003) argued, while exploring a more restrained view of multiply-engaged transitions, without any explicit underlying structure to guide such connections, this would be overwhelming for human memory to handle.

Indeed, an assumption implicit in all theories of transitions is that there is no end point for the progressive reading process. Most works have no notation for ends of a sequence the way that periods (full points) do for sentences in written language, and they (and transitions) simply end when the strip, chapter, or book itself concludes. Because of this, no limits on transitions seem to be experienced. However, experimentation has shown that readers are sensitive to chunks of scenes in sequential narratives. Gernsbacher (1985) presented readers with a series of sequential images, and found consistent agreement for where they chose to draw lines marking the boundaries of various sub-episodes. This means that people's minds are creating segments of sequential images that are not physically manifested as parts of a book, page, or layout, suggesting that comprehension does not rely on non-stop continuous linear transitions.

Structure in sequential images

If segments can be found for whole sub-episodes of sequential images, perhaps segmentation appears on a small scale as well. Take for example the sequence in figure 5.

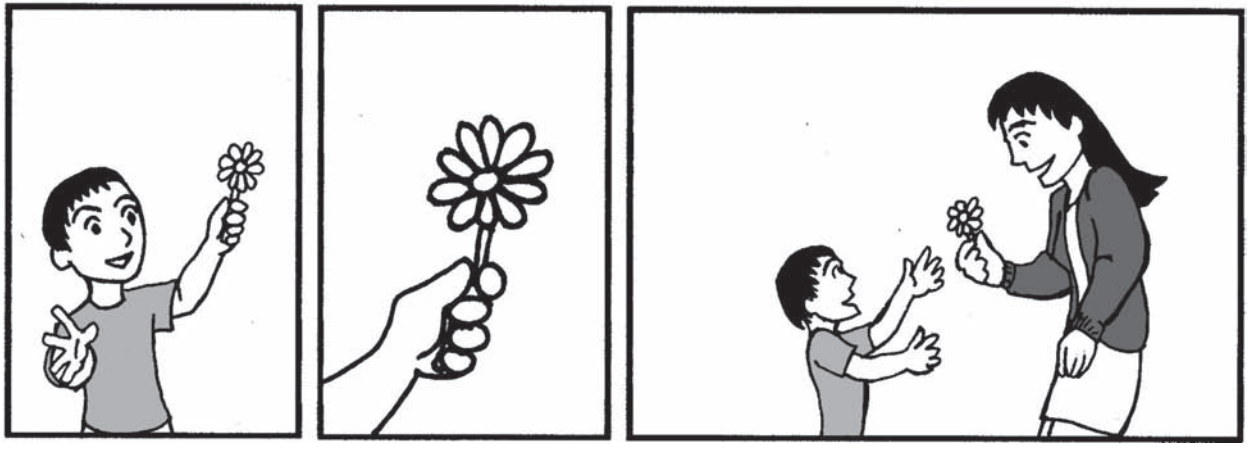

Figure 5: Zoom as a modifier.

Here, the second panel merely represents a modified view of the event state in the first panel by zooming in on the flower held by the child. Nothing about this second panel indicates that any sense of fictive time has passed, because no change has occurred to the event previously shown. If both of these initial panels represent the same state, then the progression to the third panel must take into account the unification of their contents. Indeed, by the third panel we do not forget the content of the first panel. We can confirm the single-state nature of these initial panels by observing an alternate arrangement with comparable meaning, as in figure 6.

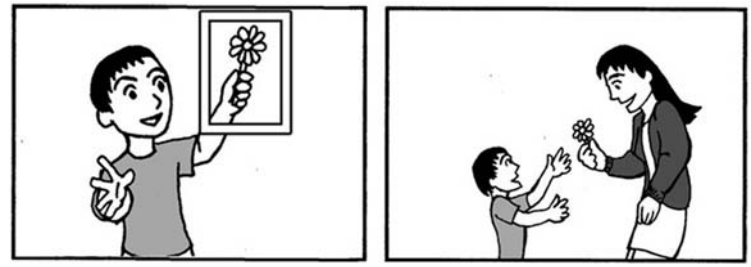

Figure 6: Inset panel serving to focus attention, like a zoom.

With the inset panel demarcating the same information as the second panel from figure 5, the spatial qualities of the panels' relations become highlighted, with no apparent time shift. Figures 5 and 6 are equal in the information they depict – only the panels serve to focus the reader's attention in different ways.

Figure 7: Temporally ambiguous initial panels.

Figure 7 shows another type of non-temporal relationship. Given its content, the second panel of this string is ambiguous as to whether it is at the same or a different state to the first panel. We know that the first and second panels represent different event states from the last panel, but the only apparent relationship between the first two panels is one of environment. We have no reason to assume that the first two panels of different people belong to a similar location. Yet, without the knowledge that those two panels belong to a singular environment, the representation of the two entities in the final panel would come as a shocking surprise. Thus, we must take into account that the union of the first two panels together must be involved in the progression to the final panel. Again, this can be confirmed by showing the exact same information through a singular panel, as in figure 8.

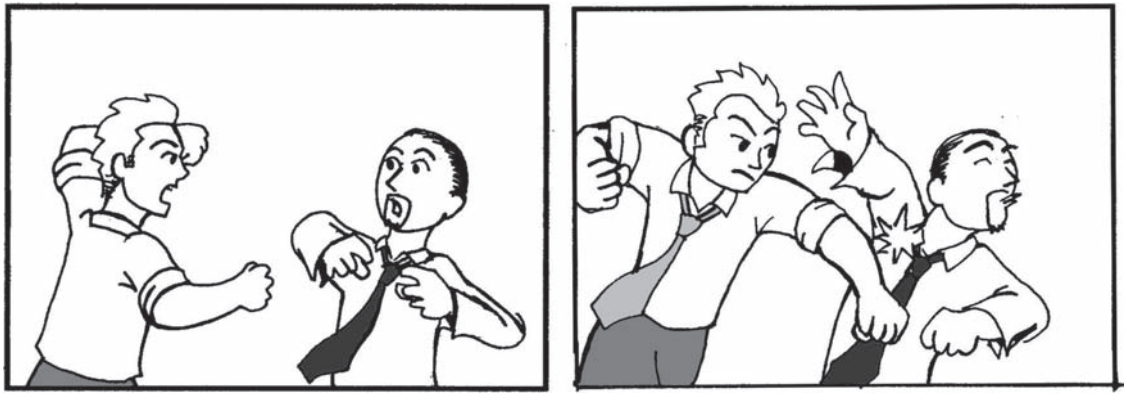

Figure 8: A single panel outset for what is accomplished in two panels in figure 7.

Because the two panels from above can equate to this singular panel, they are equivalent in their 'chunking' of information. The 'pacing' or 'narrative' might be different in two panels versus one, but they convey the same conceptual information. This means that the two-panel version must combine the initial disparate characters into a singular unit that then connects to the final panel – no matter whether they belong to the same or different states. Despite the linear reading order, this type of *understanding* is not linear, but hierarchic, and shows clearly why those panels cannot be equated to moments. Moreover, this need for grouping panels to connect to later parts shows that panel progressions do not always mimic the iconic movement of experienced events, because some sort of mental activity beyond knowledge of event states must connect non-temporal relationships.

Even more dramatically, the grouping of spatial information can work against temporal information in complex and interesting ways that undermine *any* viewpoint that linear relations alone guide the comprehension of sequential images. Take, for example, the sequence in figure 9 (Cohn 2003).

Figure 9: A structurally ambiguous sequence.

This sequence features a man lying in bed as a clock ticks away until he gets up and makes a phone call. Because the second and fourth panels both feature clocks, they must be connected somehow, as should the first and last panels that both feature a man in different stages of an action. With a transitional approach that attends to such distance connections, we would be forced to say that transitions occur between nearly every panel in this sequence, both juxtaposed and at a distance.

In fact, this sequence is ambiguous in its meaning since panels can be grouped into chunks of spatial environments as well as temporal sequences. Under one interpretation, each panel represents a separate moment in 'time', and the temporal connection of the clocks is embedded within the temporal shift between the outer panels of the man. However, a second interpretation can be created with groupings of spatial information into common environments. The first and second panels could happen in the same place at the same time – sharing a common environment – as could the final two panels. These groupings then connect in a singular shift in time.[2] These different groupings can be seen below in figure 10 using tree-structure diagrams.

2. The middle panel is fully ambiguous and could be grouped into any number of ways with the other panels. Other groupings of panels are also possible, such as a right-branching tree structure of successive moments, though not a left-branching pattern.

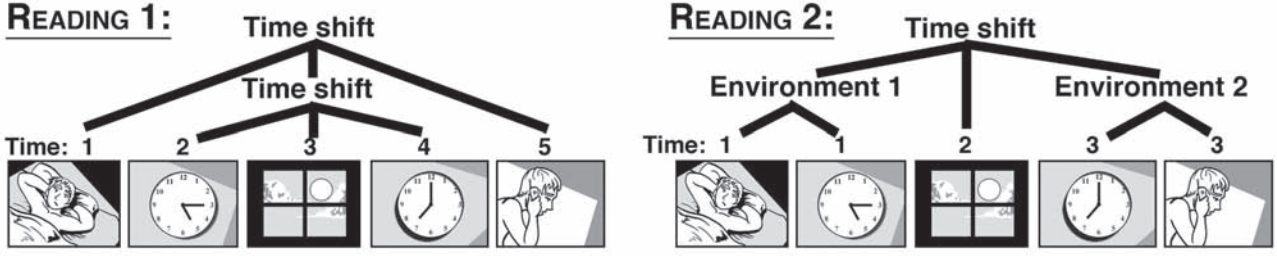

Figure 10: Varying groupings for structures of an ambiguous sequence.

Both McCloud's transitions and Groensteen's arthrology would be unable to capture the ambiguity in the sequence above. Panel transitions would merely be able to express the 'surface structure' that these panels follow (a series of subject-to-subject and aspect-to-aspect transitions), without being able to capture the differences in groupings between those panels. Arthrology would simply be able to make a statement of connection between the panels of the clocks – since they are graphically similar – but like transitions, braiding and arthrology would fail to recognize the structural ambiguity of this sequence and be unable to describe its various interpretations in any systematic way. Groensteen's approach, like McCloud's, is thus rendered insufficient for explaining the comprehension of sequential images.

These examples further demonstrate that immediately juxtaposed panels do not always represent the progression of moments of time. In all cases, panels seem to functionally divide up a conceptual space – that is additively built throughout the sequence – into units of attention (Cohn 2007). Those windowed units could narratively be whole actions, individual event states, or aspects of a spatial environment. Important to this, the meaning garnered for that 'chunking' emerges from the conceptual content of the representation itself – not from some overarching default principle like 'space = time', 'panels = moments', 'closure', or 'arthrology'.

A sequence like the ambiguous one above runs not just against the time = space views, but to any views that do not take into account deeper hierarchic connections between panels of a sequence. Much of the problem with Groensteen's theories is that they have no substantial processes to describe other than a vague sense of 'connectedness' – they make no predictions and provide no methods for meaningful analysis. Though they may allow for both local and distant relationships of graphic similarity, like panel transitions they have no notion of a hierarchic grouping of panels into chunks of a sequence.

In order for hierarchic connections to be made, they must feature an explicit set of rules and constraints for how those groupings occur. In the ambiguous sequence above in figure 9, a grouping cannot occur between panels 1 and 4 while 2 connects with 5. Such an interpretation is impossible, and there must be specific constraints – what in linguistics would be called a 'grammar' – that differentiate the acceptable groupings from the unacceptable ones. As described in Cohn (2003), the preferences for such groupings run along a scale of:

Different times >
Same time and different space/character (whole environment) >
Same time and space/character

This states that panels representing the same time and character should be grouped first, followed by panels at the same time but different characters, then finally with panels in other times. The highest nodes of a tree structure should belong to different times, the lowest to the same times. This can be made clearer by expanding the previous punching example to include a zoom as well, as in figure 11.

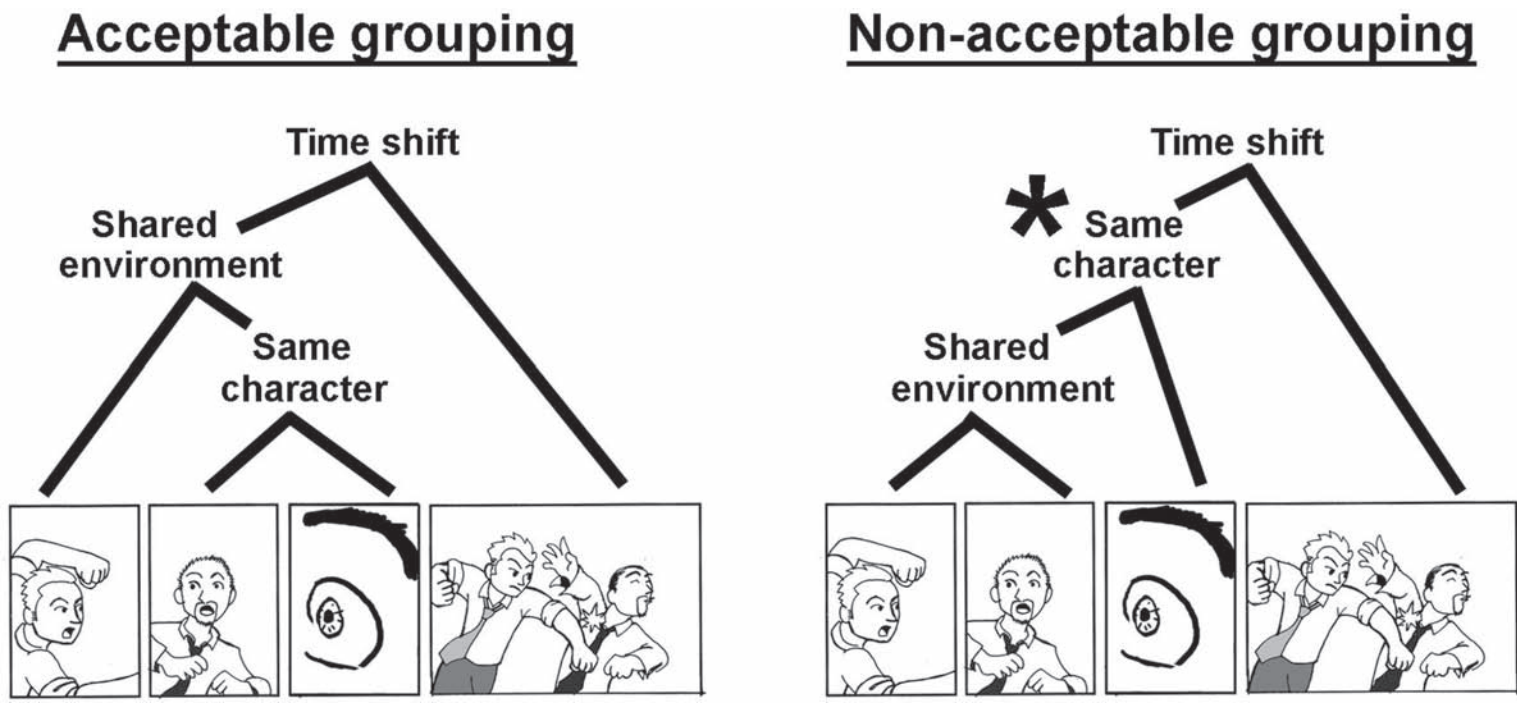

Figure 11: Allowable and disallowable groupings of time and space in sequential images.

Following the preference rules, the zoom of the second character's eye must group first with the character it modifies. From here, the grouping of this second character can unite with the first character to form a common environment, which as a whole can then unite in a shift of time. Limited by the constraints stated, the grouping in the second option would not be allowed (the asterisk indicating the problematic grouping). The common environment cannot be built first, then forced to unite with the zoom. The zoom does not modify the whole environment – it only shows a part of the second character, meaning that such a grouping would be unacceptable. This shows that the mental structures built do not have to follow the linear order to be grouped additively, since panels 2 and 3 here must come together before joining panel 1.

Where must these preferences belong? They cannot exist simply in the sequence itself, because the surface sequence (1) does not contain any overt markers of these groupings and (2) may be interpreted in a variety of ways, as in the ambiguous example above. Indeed, these preferences cannot be a part of the medium 'out in the world', but instead lie within the minds of readers and producers of sequential images.

The ramifications of such an observation are important, because they imply that the mind is actively engaged in operations directing comprehension that are completely unseen and unconscious to the reader (or author). Truly, panel transitions and arthrology are appealing as notions because the reader (or author) can directly experience them. However, the processes described here are not 'invisible meaning' in the way that McCloud talks about closure as 'filling in the gaps' (which essentially casts 'inference' as the de facto process for all juxtaposed relations). Rather, the unseen elements here are hierarchic principles of grouping provided by explicit constraints in the mind, not spontaneously emerging on the fly. These principles feature the conceptualization of time and space not in a linear sequence, but in underlying hierarchic groupings.

Conclusion

Despite its iconic nature, graphic representation does not directly mimic our naïve perceptions of temporal reality, nor are its underlying processes transparent and simple. Given that graphic representation emerges from the minds and actions of humans, it must be understood in the context of cognition – especially sequences of images, since creating a meaningful union of juxtaposed panels *must* involve mental processing beyond simple one-to-one juxtapositions. Such processing must be guided by explicit rules and constraints that involve hierarchic structures beyond linear and/or vague thematic relationships between panels. Since the 1950s, these types of hierarchic groupings have been the norm for understanding aspects of cognition for many human behaviours, from language (Chomsky 1957; 1965) and music (Lerdahl and Jackendoff 1982) to social relationships (Jackendoff 2007), vision (Marr 1982), and drawing (Willats 2005). Given such precedents, that we should find such constituencies in sequences of images should be unsurprising. As such, the study of sequential image comprehension becomes less about analysing what is 'out there' in the sequence or our conscious experience of it, but more what is inside of our own minds.

References

Barber, J. (2002), 'The Phenomenon of Multiple Dialectics in Comics Layout', Masters thesis, London: London College of Printing.

Chomsky, N. (1957), *Syntactic Structures*, The Hague: Mouton.

Chomsky, N. (1965), *Aspects of the Theory of Syntax*, Cambridge, MA: MIT Press.

Chomsky, N. (1986), *Knowledge of Language: Its Nature, Origin, and Use*, New York, NY: Praeger.

Christiansen, H.-C. (2000), 'Comics and Film: A Narrative Perspective', in A. Magnussen and H.-C. Christiansen (eds), *Comics and Culture: Analytical and Theoretical Approaches to Comics*, Copenhagen: Museum of Tusculanum Press, pp. 107–121.

Cohn, N. (2003), *Early Writings on Visual Language*, Carlsbad, CA: Emaki Productions.

Cohn, N. (2005), 'Un-Defining "Comics"', *International Journal of Comic Art*, 7:2.

Cohn, N. (2007), 'A Visual Lexicon', *Public Journal of Semiotics*, 1:1, pp. 53–84.

Cutting, J. E. (2002), 'Representing motion in a static image: Constraints and parallels in art, science, and popular culture', *Perception*, 31, pp. 1165–1194.

Dean, M. (2000), 'The Ninth Art: Traversing the Cultural Space of the American Comic Book', dissertation, Milwaukee, WI: University of Wisconsin-Milwaukee.

Gernsbacher, M. A. (1985), 'Surface Information Loss in Comprehension', *Cognitive Psychology*, 17, pp. 324–363.

Groensteen, T. (1999), *Systeme de la bande dessinée*, France: Presses Universitaires de France.

Groensteen, T. (2007), *The System of Comics* (trans. B. Beaty and N. Nguyen), Mississippi: University of Mississippi Press.

Gubern, R. (1972), *El lenguaje de los Comics*, Barcelona: Peninsula.

Harris, R. A. (1993), *The Linguistics Wars*, New York, NY: Oxford University Press.

Hünig, W. K. (1974), *Strukturen des Comic Strip*, Hildensheim: Olms.

Jackendoff, R. (1983), *Semantics and Cognition*, Cambridge, MA: MIT Press.

Jackendoff, R. (1987), *Consciousness and the Computational Mind*, Cambridge, MA: MIT Press.

Jackendoff, R. (2007), *Language, Consciousness, Culture: Essays on Mental Structure* (Jean Nicod Lectures), Cambridge, MA: MIT Press.

Kibiuishi, K. (2005), *Copper: Arcade*, http://www.boltcity.com/copper/copper_025_arcade.htm. Accessed 22 April 2005.

Koch, W. A. (1971), *Varia Semiotica*, Hildensheim: Olms.

Krafft, U. (1978), *Comics lessen*, Stuttgart: Klett-Cotta.

Lerdahl, F. and Jackendoff, R. (1982), *A Generative Theory of Tonal Music*, Cambridge, MA: MIT Press.

Marr, D. (1982), *Vision*, San Francisco, CA: Freeman.

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

McCloud, S. (1995), 'Round and Round with Scott McCloud: Interview by R.C. Harvey', *The Comics Journal*, 179, pp. 52–81.

McCloud, S. (2000), *Reinventing Comics*, New York, NY: Paradox Press.

Nakazawa, J. (2005), Development of Manga (Comic Book) Literacy in Children', in D.W. Shwalb, J. Nakazawa and B. J. Shwalb (eds), *Applied Developmental Psychology: Theory, Practice, and Research from Japan*, Greenwich, Connecticut: Information Age Publishing, pp. 23–42.

Pallenik, M. J. (1986), 'A Gunman in Town! Children Interpret a Comic Book', *Studies in the Anthropology of Visual Communication*, 3:1, pp. 38–51.

Peirce, C. S. (1931), 'Division of Signs', in C. Hartshorne and P. Weiss (eds), *Collected Papers of Charles Sanders Peirce: Volume 2: Elements of Logic*, Cambridge, MA: Harvard University Press, pp. 134–173.

Saraceni, M. (2000), 'Language Beyond Language: Comics as Verbo-Visual Texts', dissertation, Nottingham: University of Nottingham.

Saraceni, M. (2003), *The Language of Comics*, New York, NY: Routledge.

Stainbrook, E. J. (2003), 'Reading Comics: A Theoretical Analysis of Textuality and Discourse in the Comics Medium', dissertation, Pennsylvania: Indiana University of Pennsylvania.

Willats, J. (2005), *Making Sense of Children's Drawings*, Manwah, NJ: Lawrence Erlbaum.

Wilson, B. (1999), 'Becoming Japanese: Manga, Children's Drawings, and the Construction of National Character', *Visual Arts Research*, 25:2, pp. 48–60.

Wilson, B. and Wilson, M. (1987), 'Pictorial Composition and Narrative Structure: Themes and Creation of Meaning in the Drawings of Egyptian and Japanese Children', *Visual Arts Research*, 13:2, pp. 10–21.

Suggested citation

Cohn, N. (2010), 'The limits of time and transitions: challenges to theories of sequential image comprehension', *Studies in Comics* 1: 1, pp. 127–147, doi: 10.1386/stic.1.1.127/1

Contributor details

Neil Cohn researches the relationship of graphic expression – particularly the visual language used in 'comics' – with language and cognition. He has spoken on this topic internationally, and has authored

numerous articles and a book, *Early Writings on Visual Language* (Emaki Productions, 2003). As an illustrator, his graphic book (with author Thom Hartmann) *We the People: A Call to Take Back America* (CoreWay Media, 2004), addresses the pervasive influence of corporations on American government.

Neil is currently a graduate student in psychology at Tufts University, where he has taught courses on the relationship of comics to language and the mind. He received a B.A. in Asian studies from the University of California, Berkeley, studied in Japan at Tsuru University, and holds an M.A. in social science from the University of Chicago. His work can be found online at www.emaki.net.

Contact: Neil Cohn, Tufts University, Psychology Department, 490 Boston Avenue, Medford, MA 02155, USA.
E-mail: neil.cohn@tufts.edu

Visual Cultures

Edited by James Elkins

ISBN 9781841503073
Paperback | £17.95 | $25

Visual Cultures is the first study of the place of visuality and literacy in specific nations around the world, and includes authoritative, insightful essays on the value accorded to the visual and the verbal in Japan, Poland, China, Russia, Ireland, and Slovenia. The content is not only analytic, but also historical, tracing changes in the significance of visual and verbal literacy in each nation. *Visual Cultures* also raises and explores issues of national identity, and provides a wealth of information for future research.