

Beyond Speech Balloons and Thought Bubbles: The Integration of Text and Image

Neil Cohn, Ph.D.

Department of Psychology, Tufts University, Medford, MA 02155.

neilcohn@visuallanguagelab.com
www.visuallanguagelab.com

Cohn, Neil. 2013. Beyond word balloons and thought bubbles: The integration of text and image. *Semiotica*. 2013(197): 35-63

Abstract

Speech balloons and thought bubbles are among the most recognizable visual signs of the visual language used in comics. These enclosed graphic containers provide a way in which text and image can interface with each other. However, their stereotypical meanings as representing speech or thought betray much deeper semantic richness. This paper uses these graphic signs as a platform for examining the multimodal interfaces between text and image, and details four types of interfaces that characterize the connections between modalities: Inherent, Emergent, Adjoined, and Independent relationships. Each interface facilitates different levels of multimodal integration, tempered by principles of Gestalt grouping and underlying semantic features. This process allows the possibility of creating singular cohesive units of text and image that is on par with other multimodal interfaces, such as between speech and gesture.

Keywords: visual language, comics, multimodality, graphic communication, thought bubbles

1. Introduction

One of the most emblematic tropes associated with the visual language of comics is the “speech balloon,” depicting speech through a bubble that then extends back with a line towards the speaker’s mouth. Speech balloons are so associated with the medium of comics that in Italy the word “Fumetto” (“puffs of smoke”) describes both balloons and “comics” as a genre. While interesting in its own right as a symbolic representation—usually of “speech”—this sign is indicative of a greater issue: the integration of text and image into a meaningful whole. While some work has looked at cataloguing types of balloons (Forceville, Veale & Feyaerts 2010), or the meaningful relationships between the balancing of text and image in the whole of the expression (Horn 1998, McCloud 1993), little light has been cast on the structure guiding the interfaces between these two forms. Previous approaches have focused mostly on perceptual groupings, such as Arnheim’s (1974) Gestalt type perceptual ties, or Hagan’s (2007) expansion of Arnheim’s approach to integrate semantic ties. This piece will approach this issue from a structural perspective, seeking to understand the connections between the two modalities of the *written* and *visual language* found in graphic representation and the underlying properties at work in those signs.

Expanding on an earlier approach to this issue (Cohn 2003), four ways of connecting text and images can be identified:

1. *Inherent*—Relations where text and image are part of each other’s structures (Example: writing appearing in the fictive world of the drawing).
2. *Emergent*—Relations where text and image are directly interfaced with each other. (Example: Word balloons and thought bubbles).
3. *Adjoined*—Relations where text and image are integrated but not interfaced directly. (Example: Text associated to image through captions or proximity alone)
4. *Independent*—Relations where text and image are fully separate. (Example: Academic writing using “See Figure 1”).

These relationships involve varying degrees of integration between text and image in the visual-graphic modality. We now turn to addressing each type one at a time.

2. Inherent

An *Inherent* interface is the most straightforward relationship between the visual and written forms, when text occurs as some part of the image matter itself. Usually these representations manifest in banal representations, with text appearing as a “part of the fictive world” constructed in the graphics, such as words in a book someone is reading, on street signs, or any other place text might appear in the graphic world just as it appears in the real world.

In its most unique forms though, this Inherent quality of text to images may appear similar to some instances of McCloud’s (1993) Montage category which substitutes aspects of pictorial representation into written text. This sort of relation integrates the textual modality into the visual realm entirely. A *substitutive* inherency of this sort can semantically replace one modality’s signifier for that of the other (Cohn 2003). This allows one modality’s structures to merge with those of another in representation, given a shared conceptual structure. Take for instance Figure 1.



Figure 1. Substitution of text into an image

Here, the visual representation incorporates textual signs into its compositional and conceptual makeup. Though there are two distinct modalities interfacing, one is clearly structurally dominant to the other, with the secondary form becoming an inherent part of that overall structure. A more radical case of this integration happens in “visual poetry,” where words are arranged in the shape of their referent. In these cases, the visual form becomes entirely suggested by the spatial configuration of the text. Furthermore, the textual structure can become dominant to the visual, creating the reverse type of Inherent relationship. In these, a pictorial element might be inserted into a written sentence, as in Figure 2.

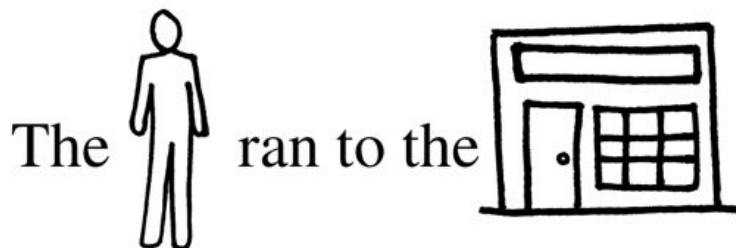


Figure 2. Substitution of images into text

This type of substitution maintains the syntax of the written form, but replaces certain signs with graphic equivalents. While such a sentence might be ungrammatical in the linguistic sense, the sheer possibility of this substitution hints that these systems are cognitively integrated. Of course, these examples feature substitution where the swapped elements remain semantically comparable, though it is not a necessity, especially in creative contexts. However, research with event-related potentials has analyzed brainwaves for incongruous versus congruous images substituted for the final words of sentences, along with incongruous sentence final words (Federmeier & Kutas 2001, Ganis, Kutas & Sereno 1996). Results showed greater amplitudes of a waveform indicating semantic incongruity—the “N400”—both for incongruous words and pictures, though each modality’s signals appeared to originate from differing places on the scalp. These results hint towards a similar cognitive processing for words and images, though with varying locations in the brain. Nevertheless, this form of substitution would be considered an Inherent relationship by incorporating both modalities as a single signal.

3. Emergent

The second degree of interaction comes from an *Emergent* relationship, where the text is bound to some visual element, though it does not “exist” in the image matter itself. Given that this relation requires a sense of connectedness, there are three components that comprise an *Emergent* interface: The *Root*, the *Carrier*, and the *Tail*, illustrated here in Figure 3.

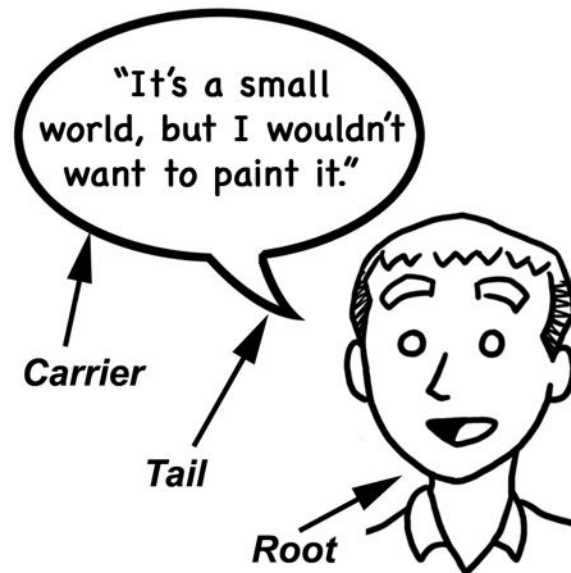


Figure 3. Parts of an Emergent interface

Because the text is not a part of the image matter itself, the Carrier is a symbolic device used to hold the text, while the Root is the visual component from which the text can be said to emerge or be integrated with. The Tail meanwhile is the index by which the two components are united (Magnussen 2000, Saraceni 2000) and, in the case of depicting speech, instances of Tails date back thousands of years (Pohl, Pope & von Nagy 2002). In many ways, the Tail facilitates the Gestalt principle of “Connectedness,” creating groupings through a physical linking element (Palmer & Rock 1994). This engenders a connection between Root and Carrier because of the grouping aspects created by the Tail, regardless of proximity to a Root. It points directionally to show where the source of the text comes from (or *to*, as in the case of Figure 3’s own diagrammatic arrows), and as such facilitates the tight integration of the text and image into a unified semantic unit. While these elements often appear overtly in demonstration, they will be understood shortly as being abstract parts that inherently make up any Emergent relationship.

In some sense, this interface creates a formalized scheme onto which various surface representations can be manifested. Carriers are familiar in the conventionalized signs of speech balloons, thought bubbles, diagrammatic boxes, and sound effects, and play a variety of functional roles in representation. Stereotypically, each of these representations stands for a particular meaning. Speech balloons show speech, thought balloons show thoughts, etc. However, representations of speech balloons especially are more promiscuous than such a simple semantic-graphic isomorphism. For example, Figure 4 uses speech balloons simply to show a purely abstract sense of pain.



Figure 4. Balloon showing pain (Riffenburg 2007). Image © Charles Riffenburg III, used with permission.

Additional uses like the first and third panels of this strip in Figure 5 from Charles Schulz’s *Peanuts* show speech balloons with non-verbal information. These balloons use a question mark and exclamation mark to express only a sense of emotion—not verbalization.



Figure 5. Balloons showing emotive information (Schulz 2004 [1950]). Image © Peanuts Worldwide, LLC. Used with permission.

Additionally, jagged borders on a balloon also can depict both loud speech, which has been shown to be highly understood by preliterate children (Yannicopoulou 2004), but can also depict sound effects. Just being “loud sounds” does not capture this distinction. In more modern mainstream comics, thought bubbles have become less popular. Instead, narrative captions with first-person dialogue such as those in comics like the popular book *Watchmen* have provided the same window into character’s thoughts, just in a different surface representation. Both carry the same meaning, with different representations.

Instances like these underscore the need for a more expansive semantic theory for these signs. While straightforward meanings like speech and thoughts must be involved, they cannot be the whole picture (so to speak). A good place to start addressing these issues can be with who is aware of the contents of the Carriers.

On the whole, despite the variety of roles that they play, all Carriers function similarly on an abstract level for encapsulating text, which often emerges from a Root in the graphics. However, neither the writing nor the Carriers conventionally exist in the constructed world of the

narrative, unless there is some violation of the characters' obliviousness of them as “invisible and intangible objects” not merely representing sound, such as if the character grabbed the text or balloon, as in Figure 6.



Figure 6. Violation of awareness of balloons.

However, without such a violation, the variability of a Carrier’s function revolves around the degree of “awareness” of the content of the Carrier by entities in the image. For example, while speech balloons and sound effects contain sound utterances “heard” by other entities in the representation, the contents of “thought balloons” are accessible only to the Root connected to the bubbles (Dean 2000, Murray 1997). These semantic roles that Carriers play can be clarified by decomposing these properties into a set of underlying functional primitives motivated by such awareness. If the contents of a Carrier can possibly be “recognized” by the Root (as in both word balloons and thought bubbles), it is motivated by *Root Awareness* [+RA]. If other surrounding “non-Root” entities within the scene have awareness of the contents of a Carrier, they have *Adjacent Awareness* [+AA] (as in word balloons, but not thought bubbles). With both positive and negative values, these primitives can now be seen to apply in varying combinations across all types of Carriers.

1. *Root Awareness* [±RA]—Recognition of the contents of a Carrier by the entity it is interfaced with (the Root).
2. *Adjacent Awareness* [±AA]—Recognition of the contents of a Carrier by entities other than the one it is interfaced with (all other non-Root entities).

If a Carrier features both Root [+RA] and Adjacent Awareness [+AA], all entities in the scene recognize their content. These *Public* Carriers most often manifest as speech balloons, which symbolize sounds in the depicted world produced by a source that is aware of it, and understood by any entities in the surrounding environment. Of course, this presumes that all the entities within the scene have awareness, most prominently the Root.

Public Carriers are contrasted by sound effects that come from a non-sentient Root, such as a gun, or some other non-living object. In these cases, the Root cannot have an awareness of its emergent sound because it has *no awareness*, yielding a *Non-Sentient Carrier* [-RA, +AA]. Of course if the non-sentient Root were anthropomorphized—such as a cartoon gun that talks and shouts “Bang!” instead of just making the noise, it would then shift into a Public type of Carrier. This could be accomplished with a surface representation of either a word balloon or a sound effect.

The complete inverse of this type would be a Carrier recognized by the Root [+RA], but not by surrounding entities [-AA]. These *Private* Carriers most often appear as thought balloons,

which represent the internal monologue of individual characters. In some cases, Private Carriers appear for animals as thought balloons to depict their “speech” that cannot be heard by humans—as in Snoopy’s dialogue in *Peanuts*. However, Private Carriers might also appear as “speech balloons” with unutterable symbols like “!” or “?” in them, or alternatively as narrative captions expressing the thoughts of a character in the panel. Provided that the Root is depicted visually in the scene, this type of first person narration can be considered merely as an aesthetically different form of a thought balloon. These are all surface representations of the same underlying semantic features. This emphasizes the point that this analysis is not merely positing underlying fields of meaning for certain depictions, but rather that there are tacit primitives beneath all the varying *surface* forms of Carriers that are employed throughout visual and textual interfacing.

Lastly, there are Carriers that have content unrecognized by any entities in the scene, including the Root [-RA, -AA]. These *Satellite* Carriers are recognized only by the reader, and often appear as diagrammatic boxes with Tails pointing to things that they discuss (for example, in instruction manuals). This type of Emergent text is recognized neither by the Root or the surrounding entity, but only by the reader. Here, the Root may not “produce” this text, as in the other types of Emergent Carriers, though it is still bound to the content of the image through a Tail, upholding the tripartite division of Emergent relations. Because of its functionally negative value for awareness, this type of Carrier is closer to the next level of interface beyond Emergent relations than any of the others.

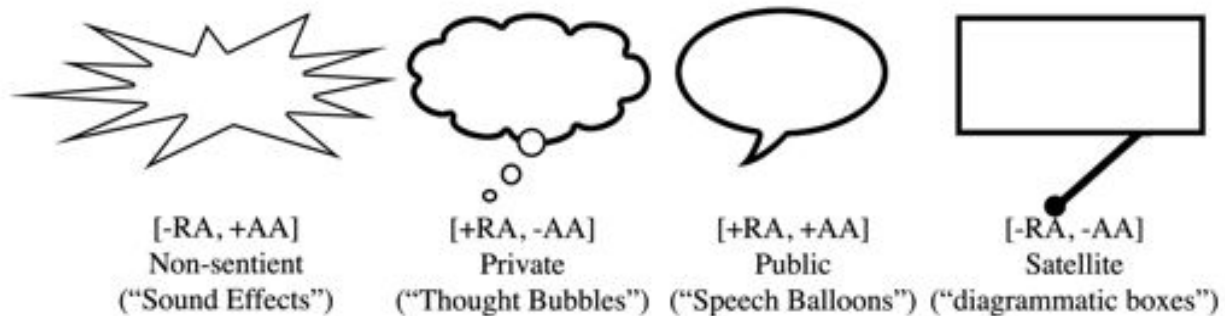


Figure 7. Prototypical representations of types of Carriers

These four types of Carriers are summarized along with their prototypical representations in Figure 7. While combinations of these functional primitives prototypically map to certain surface depictions (ex. Private → Thought Bubbles), there can be mismatches between underlying features and surface representation. For instance, if a character in a scene turns directly to the reader and speaks in a word balloon as an aside, it functionally becomes a [-Adjacent Awareness], because the other entities in the scene cannot recognize it. Or, the “speech” balloon in Figure 4 containing only graphic stars and pins pointing to the person’s back acts as a Satellite symbolizing the pain they feel but not “speech.” Thus, previous analyses of these phenomena as equivalent to direct or indirect speech (e.g. Saraceni 2000: 207) are merely describing the surface results of a more complex interaction of features.

3.1 Tails and Indexicality

The Tail functions to bind the Carrier with the Root, and as such marks where the flow of connection travels from one to the other. This can be highlighted through some creative representations, for instance, when the Tail itself becomes a Root, as in Figure 8.

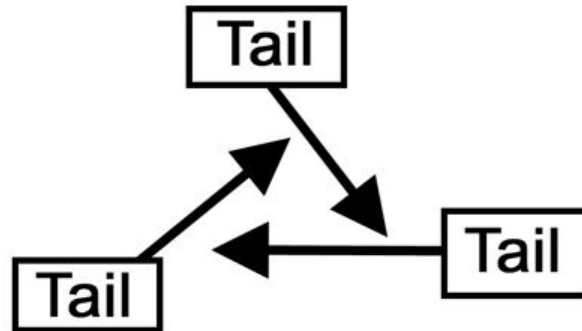


Figure 8. Recursive Tail-Root indexicality

While this Root-Tail-Carrier relationship might remain atypical, it exemplifies the centrality of the Tail as the interfacing agent in the Emergent relationship. Because the Tail here picks out the Root, when indexing another Tail it allows for a recursive loop of indication. This indexical power of the Tail is especially salient when the Root remains unseen, as in Figure 9.

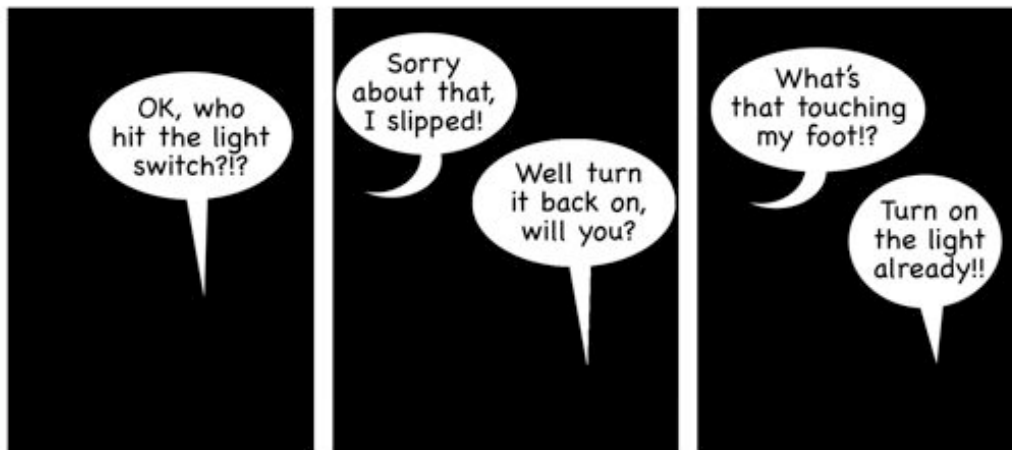


Figure 9. Unseen Roots

These panels show a scene where darkness conceals the visual appearances of the characters. Since the Roots are not visible, the Tails indicate how many entities are evident in the scene. In the first panel, only one Carrier and Tail are shown, indicating that only one entity is “depicted.” In the second two panels, since the Carriers have both Tails pointing in opposite directions, the reader understands that there are two characters present. If both Tails pointed towards one direction, only the presence of one entity would be discerned. Moreover, the curved Tails on the upper balloons indicate that character might even be outside the borders of the panel. In this way, Tails are bound to their Roots, indexing their existence even when they are not shown.

Public Carriers seem to have at least minimal restrictions for Tails, guided by their indexical value. Since they are often meant to represent speech, Tails should point in the direction of a speaker's mouth as opposed to the general vicinity of a speaker, and informal traditions passed on by letterers of comics have claimed that Tails should optimally end 50 to 60% of the distance between the Root's head and the Carrier (Piekos 2009). While there appear to be some allowances for variability in their directions, Tail's varying radically from this might seem a bit awkward. For instance, despite the knowledge that it should represent speech, the problematic third example in Figure 10 makes the balloon to seem to emerge from the speaker's feet.

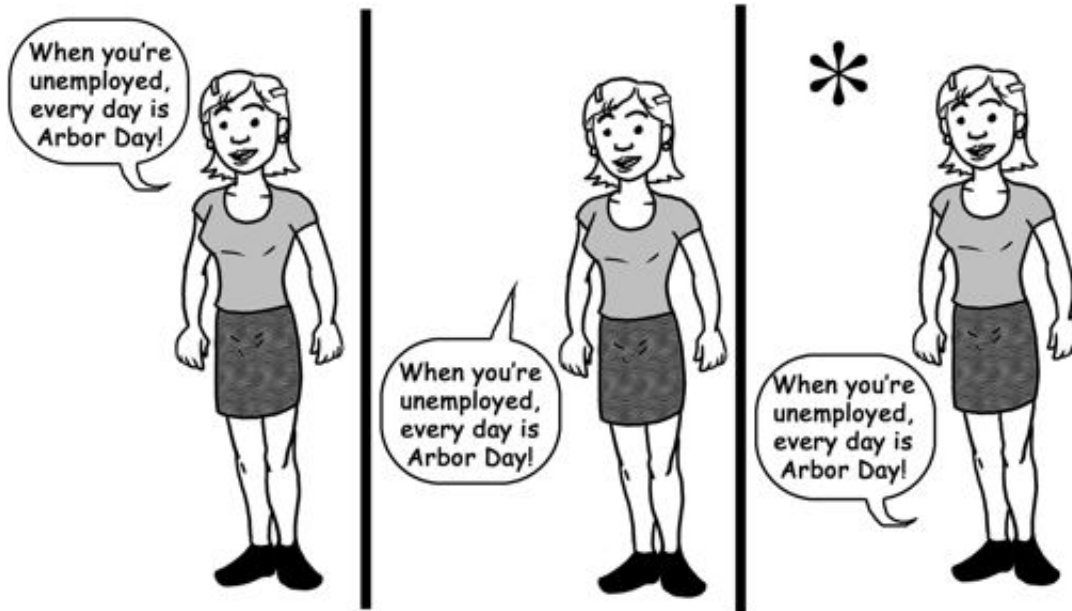


Figure 10. Indexicality of Public Tails

To common sense, speech balloons emerging from non-speaking objects (like feet) should seem somewhat awkward. However, in graphic form any object can plausibly “speak” because the language is not bound by “real” physical properties. In this case, the direction of the Public Tail assigns the Root role to the feet, while subjugating the person they belong to into an adjacent role. Private Carriers have a similar restriction. The Tail is expected to go towards the head of a person, though other non-thinking objects can take such Carriers without problem as in Figure 11.

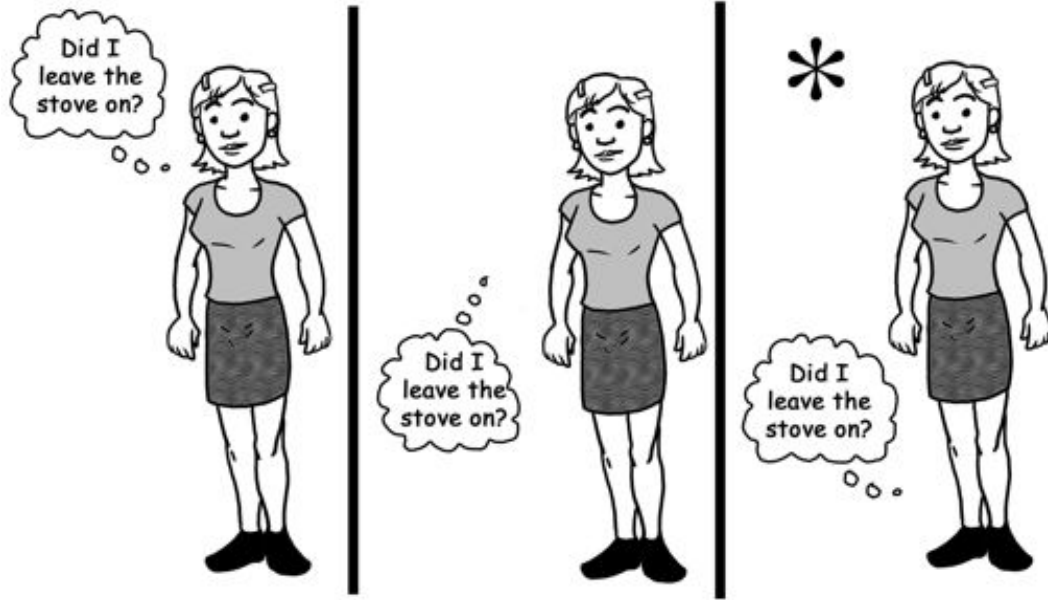
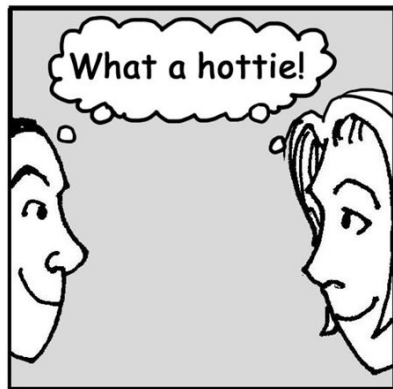


Figure 11. Indexicality of Private Tails

The import of Tails’ indexical value emerges further when a Carrier has more than one Root. Readers do not interpret the content of one balloon being “shared” by several Roots, but rather that multiple Roots say or think the *same thing at the same time*. This is especially salient with regards to Private Carriers, as in Figure 12, since each Root acts as an adjacent entity to the other. Neither knows the content of each other’s thoughts, even though they are represented in the same balloon!



a. Positive shared thoughts



b. Adversarial shared thoughts

Figure 12. Adjacent Awareness restrictions in multiple Tails

Thus, while the symbolic Carrier might encode the semantic primitives (in this case [+RA, -AA]), it is the indexical Tail that distributes these primitives to their Root individually. However, the restriction of acute direction for the Tail does not exist in Emergent relations where the Root has no awareness, such as [-RA, -AA]. The Tail is only restricted by how closely the semantics of the text within the Carrier aligns to the Root.

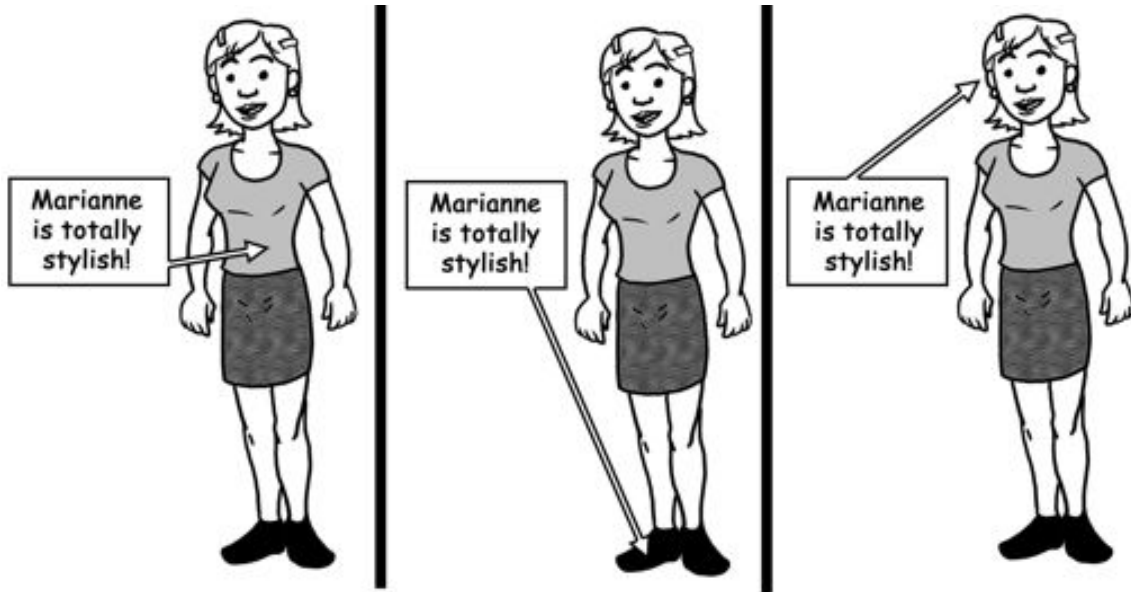


Figure 13. Indexicality of Satellite Tails

In Figure 13, the Tail could hypothetically point to any part of Marianne’s body because the text only refers to her as a whole, though it seems to draw attention to whatever piece of clothing/etc. it is pointing towards as being the motivation for calling her “stylish.” However, if the text of indication is specific, as in Figure 14, and conveys information smaller than the whole depiction (in this case smaller than “person”), variation of Tail direction can create dissonance in meaning.

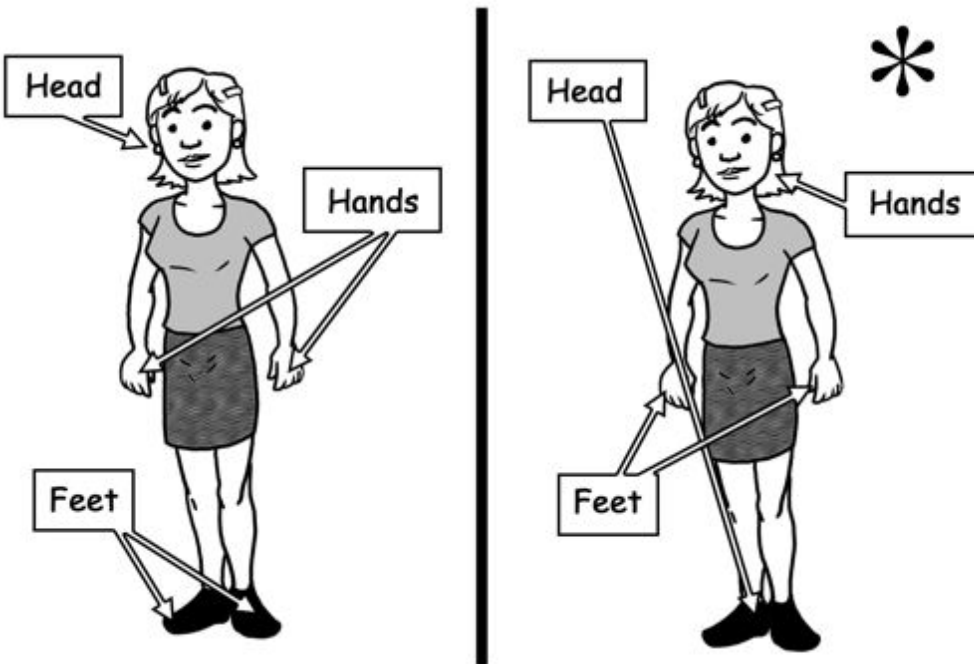


Figure 14. Indexicality of Satellite Tails with specific mappings

Just as two Tails can occur with Private and Public Carriers, Satellite Tails can also be used on Carriers of other types. For example, note this type of representation in Figure 15. Here, the balloon acts as a Public Carrier coming from the Root of the person, though with the addition of the arrow, the toaster becomes a second Root that holds a Satellite relationship with the Carrier. In this case, the Satellite style Tail motivates a second Emergent relationship for the Carrier. Both Tails act as arrows indicating their relative relationship with a Root, but in assigning the direction of indication relative to the Carrier it emerges *from* the person but *towards* the toaster.



Figure 15. Indexicality split between two types of Tails

This seems to imply that a positive Root Awareness [+RA] semantically places the Tail’s indexical value more with the Root, while its absence places the restricting factor on material inside the Carrier. Essentially, [+RA] makes the Carrier’s contents seem to *emerge from* the Root *into* the Carrier, while [-RA] makes the contents seem *directed at* the Root *from* the Carrier. However, [+RA] is not the only factor involved, as [-RA, +AA] relations retain an indexical direction going away from the Root (as in “sound effects”). We can map these relations in a grid as shown in Figure 16, which indicates that indexicality is directed towards the Root only in the absence of both Root and Adjacent Awareness. In other words: where there is no entity awareness at all. This directionality of connectedness becomes even more important when factoring in the roles of Roots and Carriers in representation.

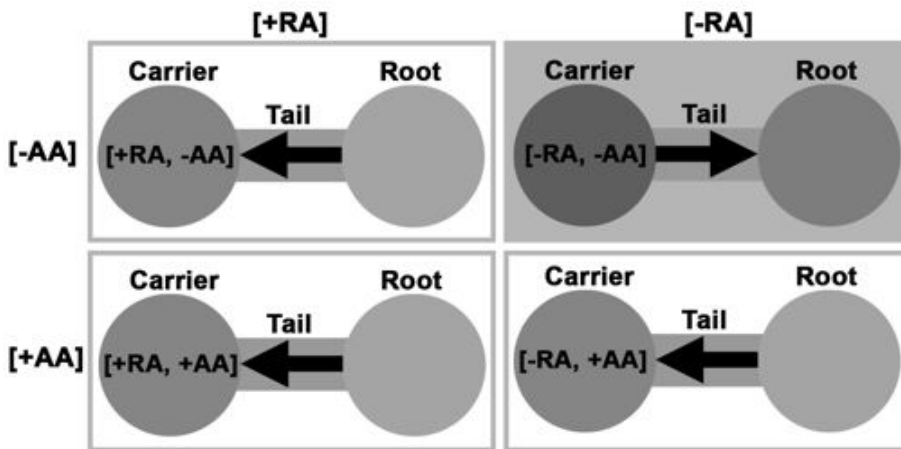


Figure 16. Direction of Indexicality for all Carrier relations

3.2 Structural Primitives and Two Planes of Representation

Another aspect of this indexical relationship relates back to the semantics of the Root. Because all Roots attached to [+RA] Carriers must possess some sort of “awareness,” connecting a [+RA] Carrier to any object will immediately convert its own semantics to a [+ANIMATE] feature. That is, no matter the visual object, the juxtaposition of a [+RA] Carrier will imbue it with animacy, as in the “talking shoe” example earlier.

This knowledge of “awareness” relates to the concepts of “Theory of Mind” that have emerged in the past several decades of psychological research (Premack & Woodruff 1978). The Theory of Mind focuses on the cognitive ability that humans can recognize the existence of mental states in other people. People know that others have thoughts, and interact with them on the basis of that knowledge. In the case of the semantic primitives for Carriers, “Adjacent Awareness” reflects a Theory of Mind quite clearly, and helps motivate “thought balloons.” Indeed, children as young as 3 years old have been shown to be able to understand that thought balloons contain intangible mental contents (Wellman, Hollander & Schult 1996), though preliterate preschool aged children only recognize the difference between speech and thought balloons as directly showing speech or thoughts at chance (Yannicopoulou 2004). This representation has been reversely put to use in studies where thought balloons aid autistic children to recognize such mental states in others (Wellman, et al. 2002).

Furthermore, Carriers are not necessarily an explicit holder of textual information only, but can also hold visual matter as well (Bongco 2000, Dean 2000), as in Figure 17. This sort of usage of Carriers does not use text at all, though it takes characteristics of a multimodal interface.



a. Emotive content in balloon



b. Shared image content in balloon

Figure 17. Images in Carrier content

Beyond the expression of emotion, as in the above examples, this property of Carriers for holding visual information can be exploited in several interesting ways. For instance, it can be used to show the recursive power of visual sequences. Recursion has been emphasized as central feature of human language and cognition in that it allows for a grammar to generate infinitely long sequences from finite means (Chomsky 1995, Hauser, Chomsky & Fitch 2002). Recursion exemplifies the creative capacity of language in that it disallows the possibility of memorizing grammar by rote combinations, because any sentence can potentially be augmented to become longer, to an unlimited length. This property can be seen simply by adding the phrase “X says/thinks/suspects/etc that...” to the beginning of any sentence (Jackendoff 1994: 14):

Jared likes watching cartoons.

Christina says that Jared likes watching cartoons.

Beau says that Christina says that Jared likes watching cartoons.

This addition could go on infinitely. It even can fold in upon itself provided that there is only one person named “Jared” in the following (peculiarities for not using a pronoun notwithstanding):

Jared₁ says that Beau says that Christina says that Jared₁ likes watching cartoons.

Because a Carrier of a speech or thought balloon can carry out the same function, this recursive power can easily be applied visually if the Carrier is manifest as a panel itself. The use of this type of a Carrier as a panel achieves the same affect as “X says that...”, and can apply recursively to an unlimited extent, as in Figure 18. In this example, the Adjacent Awareness tapped into by the Public Carrier is only accessible to whoever is in the balloon with the Root. This creates a cascading effect such that Roots become Adjacent entities as one moves rightward in the sequence, with each potentially knowing the contents of all the subsequent Carriers.



Figure 18. Recursive Carriers as panels

Not only does this demonstrate recursion in sequential images, but it also reveals a great deal about the relationship of Carriers to panels. To this end, Carriers can be seen as micro-sized panels, similar to Inset panels that are set within another “dominant” panel (Cohn 2003, 2007). Indeed, the above sequence could easily be drawn as a series of panels embedded into larger panels (the first *inside* the second *inside* the third...etc), instead of unfurled in separate panels linearly. In this way, both Carriers and panels (and arguably pages) can be viewed as framing devices that divide the semantic components of this visual language into attention-drawing units in a variety of interacting ways. Indeed, as will be discussed further for Carriers, panels can also feature a “null value” in representation, as in the case of borderless panels, and collage-styled representation.

The sort of encapsulation featured by framing devices like Carriers and panel borders can be described by separating the depiction into two planes: an inner one of content—the Representational Plane (RP)—and an outer Framing Plane (FP). The Representational Plane contains all the “visual content” for the sequence, while the Framing Plane contains devices such as Carriers, panel borders, and text (with the exception of Inherent text which is part of the RP), and has emerged through conventions over time in the visual languages of much of the world to

further order and compose sequences of images and their interfacing with text. These split planes are illustrated in Figure 19.

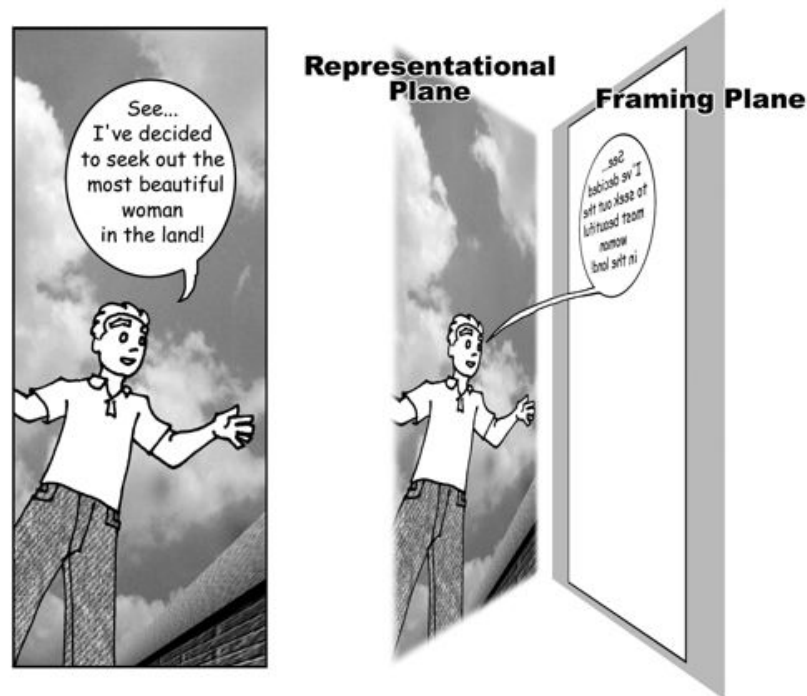


Figure 19. Split planes of representation in panels

Abbott (1986: 156) observed this two-layered distinction as well, dividing textual phenomena into three groups: narration, dialogue, and sound effects. While both narration and dialogue are not actually part of the image matter, he claims that sound effects carry over to the RP and “function in an intermediate position between the literary and the pictorial” (Abbott 1986: 156). This view perceives sound effects as more “pictorial” in nature, due to their frequent lack of Carriers and more visually engaging composition than text in speech balloons or narrative boxes. Placing sound effects as part of the RP though is problematic because they still use an Emergent relationship, not an Inherent one. Despite the visual artistry that they might take, they are not “in” the narrative world. Rather, any Emergent relation bridges this separation between planes, because the Tail (either manifest or abstract) provides the link between the two levels.

This link still does not place *Emergent* text as part of the RP though. It merely reflects the binding connection between the two planes achieved (abstractly) by the Tail, which in representation belongs in the Framing Plane. Indeed, while the FP integrates the Carrier to the RP through the *Emergent* interface, the primitives underlying that Carrier guide how its content interacts meaningfully to the content of the RP.

Japanese theorist Fusanosuke Natsume elaborates on this split plane juxtaposition, describing that the RP itself has both “positive” and “negative” characteristics, roughly aligning to figures versus backgrounds (Natsume 1997). The “gutter”—the space between panels—he implies, is part of the outer component of the image’s frame. In that both of these levels are visual, like the previous “recognition” of the speech balloons, characters can interact with the panel borders as “physical objects” within the bounds of their Representational Plane. When a character engages the panel borders, they engage the space of the frame, pulling the positive figures away from the negative background space. However, this distinction exists only in terms

of the RP levels of Natsume’s breakdown, and does not entirely address the level of the frame itself. Rather than one of these planes being brought into the sphere of the other when a character interacts with panel borders, these instances pinch the two planes together. Indeed, this does not bring panel borders inside the constructed world of the visual content, nor does it pull the character outside of the constructed world within the panel. Instead, it is an admittance of the dual planes, playing off of their mutual visual form. However, the transference of the Framing Plane into the Representational Plane could conceivably happen by enclosing the panel borders as part of the constructed Representational “world” and thus excluding them from being part of a second plane as demonstrated by Figure 20.



Figure 20. Interaction of characters with Framing Plane

Here, by breaking out of the panels, and then referencing them as being enclosing (“I’m out!”), the depiction pulls the panels themselves into the Representational Plane, though the Carriers remain in the Framing Plane, diagrammed in Figure 21.

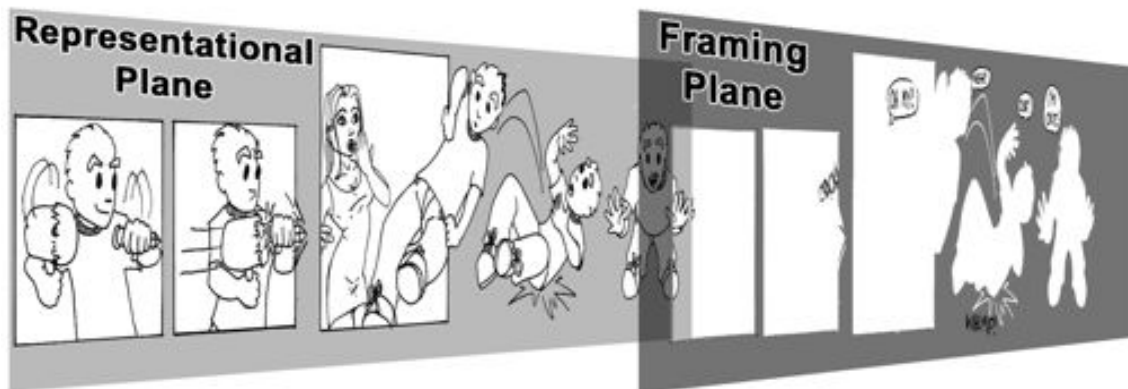


Figure 21. Interactions between planes

In this case, when the character emerges into borderless panels, the panel borders enter the images’ representational makeup and not a framing device. As such, the panel borders exist in the RP, and not the FP. Note also in the previous example with characters’ “recognition” of speech balloons that it did not change the value of the semantic primitives underlying the Carriers. Rather, their semantics remained the same—as part of the FP—while the instance merely represented a pinching of the FP and RP. The characters still had awareness of the Carrier’s content in the same semantic way, but the change came in suddenly becoming part of their conceptual world, rather than an extradiagetic device outside of the RP.

Just as panel borders might not need to be shown, not all Emergent relationships feature visually represented components. In some cases, only a simple line will appear for a Tail without any depicted Carrier (Figure 22a), or often sound effects will appear without any overt Carrier or Tail at all, appearing only visually juxtaposed from the Root. Also, sometimes the text itself might take the form of a Tail, to visually indicate the interface through its own composition (Figure 22b). This will be deemed as *Compositionally Enclosed (CE)*.



a. Null Carrier (on both Carrier and Panel)
Figure 22. Null Carriers



b. Null Carrier with CE Tail

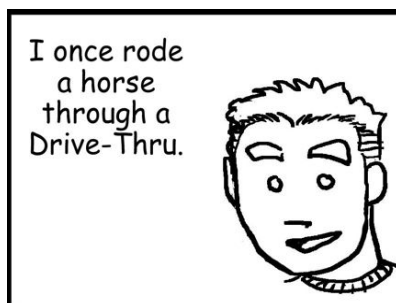
From all this, we can posit two further primitives for Emergent Carriers. To complement the functional primitives discussed previously, a set of structural primitives can also be added:

Emergent Carriers
Functional Primitives
 [± Root Awareness]
 [± Adjacent Awareness]

Structural Primitives
 [± Carrier Representation]
 [± Tail Representation]

Thus, there are four variables that control the way Carriers are employed functionally and represented visually. Compositionally Enclosed Tails are considered [-CR, -TR], yet “covertly” positive, meaning that their Tail is subsumed by the depiction of the text itself. In other words, indexicality becomes an attribute given to the representation of text rather than the Carrier, though fulfilling the properties of the Carrier.

However, there may be restrictions on the null value for structural representations. For instance, though it is recognized as Emergent, speech with both null values for Carrier and Tail seem somewhat more removed from the Root than if they were represented visually, whether or not enclosed by a panel, as in the examples in Figure 23.



a. Inside panel
Figure 23. Null Carriers and Tails



b. Outside panel

Though it is capable of being connected to its Root, the text here in Figure 23 seem somewhat more distant than if they had a represented Tail. Nevertheless, it is recognized that the invisible Carrier holding the text would be Public [+RA, +AA], rather than Private [+RA, -AA] despite no graphic indication as such. This “looseness” in connecting the text to the Root most likely stems from a lack of cohesive integration between the two forms, shirking the Gestalt principle of Connectedness, and relying only on Proximity (Wertheimer 1923)—grouping created by nearness of physical distance—or the “Common Region”—grouping created by elements sharing a common space (Palmer 1992)—created by the enclosure of the panel borders.

Furthermore, the Root can also be given a $[\pm]$ potential in representation, as it is easily removed from the visual representation with the relation upheld, as in Figure 24.



a. Null Root (outside panel)



b. Null Carrier and Root with CE Tail

Figure 24. Null Carriers

In these instances, the Emergent text serves as a referential index for the unseen Root—even if *none* of the Emergent relation structures appear visually overt, as in the right example. Imagine a scene where characters hear continual sounds off in the distance, which are not framed by any sort of Carrier. Because of the distance of the sound within the narrative, Tails might not be employed, and the Root of expression is not shown either. However, the reader *implicitly understands* that this Emergent text does have a source, and thus must be connected in some manner of interface. Because the Root might not be drawn yet still inferred, it serves as a testament to the tripartite relation of the Emergent interface.

If Carriers and Tails have null values in representation, it applies directly to their depiction on the Framing Plane. On the other side, the Representational Plane holds the consistency for whether or not the Root is shown. This division further reflects the interfacing nature of Emergent relations. Since it always contains three parts (the Root, the Carrier contents, and the Tail), we can say that components can be “bundled” together: the visual integration of text and image form a unified semantic unit, even if components are not represented. This Principle of Bundling most often occurs through some apparent visual interface, such as the overt Tail relation in an Emergent interface, or the covert inherency of the textual to the visual in Inherent relations. Both of these types of interactions are always bundled.

3.3 Multimodal units

Several conceptions have been proposed that express the unification of multiple modalities into a single unified whole. For example, Mitchell (1986) has referred specifically to the synthesis of the verbal and visual as “imagetexts,” a term that has been growing especially

amongst literary scholars of comics to the point that it has been chosen as the name for a journal. From a psychological perspective, Clark (1996) has referred to the integration of multimodal communication into single units as “composite signals,” a term he has applied to the union of spoken utterances with manual gestures, though which can also apply to text-image relations. Related, McNeill (1992) has noted that gestures appear roughly once per spoken clause, and also emphasizes the common conceptual origins of both modalities from a “Growth Point.” By emerging together, these “concurrent gestures” coincide with the spoken utterance, creating dual routes for a single overarching meaning.

Similarly, the combination of text and image can concatenate meaning in the graphic form. Because the visual domain does not rely on concurrent messages in *time*, integration must occur in *space*—such as through panel borders or Emergent interfaces. This spatial integration creates composite signals in the graphic domain comparable to the temporal integration of speech and gesture. However, if text and image are not conjoined graphically in this way, such as text distant from an image or layered on top of panels, the graphic integration of these forms does not occur, and connection relies solely on the semantics of each modality.

As such, bundled panels will be referred to as *composite signals*, to reflect their integration as semantic units. However, “unbundled” phenomena, which still must have semantic interactions on a larger scale, could be referred to as *composite syntagms*, reflecting both the combination of multiple composite signals and the integration of unbundled text and image, to which we now turn.

4. Adjoined

One step further removed than Emergent interfaces would be that of an *Adjoined* interaction of text and image, in which text is visually integrated, though not directly connected to some element within the RP, as in Emergent interfaces. This interface is found in most forms of narrative captions that hover above or overlay the visuals, yet are not directly connected to the visual representation. While Adjoined relations use Carriers—especially narrative captions—they do not employ Tails, because there is no interface to the RP as in Emergent relations. Likewise, Adjoined Carriers cannot be “recognized” by any entities in the scene, because they do not contain any Root, and thus cannot appear to any adjacent entities either. Thus, starting off, they are far more constrained in their primitives to [-RA, -AA; ±CR, -TR].

Adjoined Carriers

Functional Primitives

[- Root Awareness]

[- Adjacent Awareness]

Structural Primitives

[± Carrier Representation]

[- Tail Representation]

This means that “narrative captions” function as Satellites with no Tails [-RA,-AA] since the contents of Adjoined Carriers are only “recognized” by the reader, and feature no connection to the RP because of the lack of Tails. However, as the only present feature of Adjoined interfaces, they can vary in their visual depiction of Carriers. To this end, unlike Emergent interfaces, Adjoined Carriers feature no link from the Framing Plane to any sort of Root, though they might be connected via the visual inclusion of the panel borders. This can even apply if the Adjoined text represents a “voice over” of an entity’s narration—provided that that the “narrative captions” are not actually Private Carriers, and thereby have no connection to the panel content

outside of semantics because they are not *directly* interfaced with the RP as in Emergent relations.

As such, Adjoined relations are not always bundled, though they can be. In Gestalt terms, these Carriers would not be bound by Connectedness (because of no Tail), though may obey Common Region by being enclosed in the same panel borders as the RP content. Semantic connection of the two modalities aside, bundling does not always occur between the two forms, despite their visual integration. Though it may seem absolute as a unit (Abbott 1986, Bongco 2000), visually, the panel borders are not a rigid “bundling device,” as the text can easily separate itself into a different track of reading unbundled from the visuals as in Figure 25.

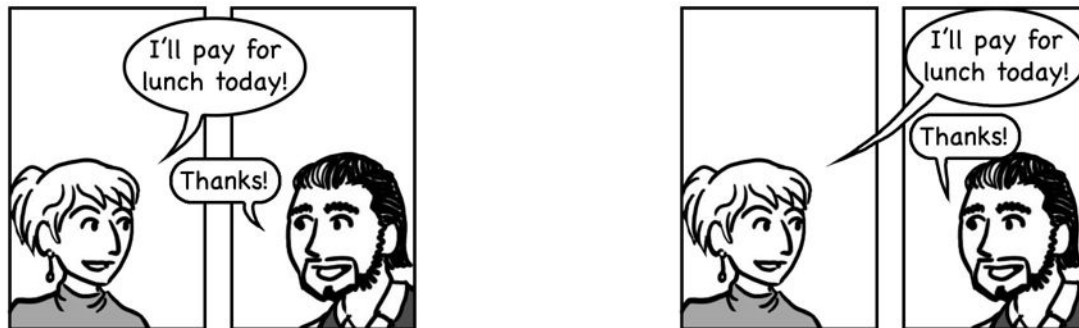


a. Bundled

b. Non-Bundled

Figure 25. Bundled versus unbundled Satellite Carriers

By separating Carriers from the confines of the panel borders, Adjoined text can take on a reading track separate to that of the visual sequence. This forgoes the bundling that occurs at the panel unit level, yet still allows semantic integration for the broader whole of the reading. However, while this disparate tracking separates the bundling of Adjoined relations, its effect on the bundling of Emergent interfaces is variable on its content. For example, if the panels feature two different characters, then Carriers can overlie both on the gutter (Figure 26a) and even from a first panel into a second (Figure 26b).



a. Overlay on gutter

b. Crossing panels

Figure 26.

In both of these cases, the bundling of the panel is weakened compared to the bundling by the Tails. In Gestalt terms, Common Region loses out to Connectedness. This is partially

facilitated because the content of the panels features two characters of a common shared environment. Each panel is windowing only a portion of the overall scene, and thus the panel borders are simply directing attention to each characters portion of the environment (Cohn 2007). Essentially, it acts like there is an invisible “mega-panel” combining these two panels together, such that the individual frames make little impact. However, if a single character is repeated across panels, this “mega-panel” disappears, and the overlay becomes infelicitous (Figure 27).



Figure 27. Infelicitous crossing of panels

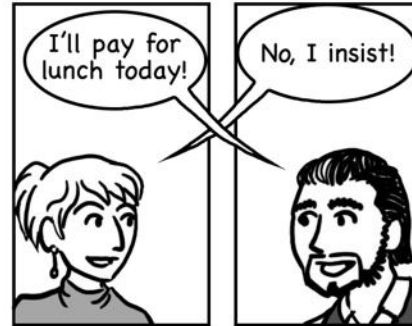


Figure 28. Crossing Tails

Here, the Carrier from the Root in one panel crosses over to that of the subsequent panel, which features a later state of the same Root. This becomes problematic, likely because Common Region here becomes a more salient grouping feature since each panel demands that it be a bundled unit. Interestingly, even when separate characters are featured in each panel, if both Carriers cross panel borders, it does become problematic, as in Figure 28. Here, both Common Region and Connectedness are pitted against each other in a way that does not allow either one to “win out.” Notice that an Adjoined relationship has no problems with this overlay, regardless of the content of the panels (Figure 29).



a. Caption over separate characters
Figure 29.



b. Caption over same character

Because they do not mandate bundling through a tight relationship the way that Emergent relations do, Adjoined relations involve a more freed relationship to the content of the images. Bundling is left almost wholly to the compositional encapsulation by panels, or else they rely on mere proximity and/or the semantic content of the words and images to integrate them into a holistic meaning.

5. Independent

Finally, an *Independent* relationship separates visual and written forms the most, featuring no visual connection between the two modalities at all, despite a semantic association. In other words, each modality is kept separate with no visual connections whatsoever.

Such interactions are often used as standard in modern writing practices, with the typical non-visually integrative “See Figure 1” type interface used for every example in this paper. However, an Independent relationship is only capable if there is some sort of connecting element between the two forms. Essentially, the text “See Figure 1” serves as the indexical interface for the two modalities. This is starkly different, for example, from newspapers or magazines, which may show an image illustratively (with an Independent or even Adjoined caption), but then only tacitly refer to it from the text of the article. Here, there is no actual interface between the article’s text and image, but only a general indexing made by the recognition that the image pertains to the article.¹

Because this interface contains no visual interaction at all, there is no bundling capable in Independent relations, which is why these relations are so difficult to integrate. Independent relations thus have no capacity to create composite signals—they are not concatenated semantic units at all and have neither temporal nor spatial concurrence. Nonetheless, Independent interfaces are the standard in academic and technical writing (such as this paper), likely holdovers from less technologically proficient eras of printing. Perhaps as the technology becomes more accepted, this segregated view of modalities will dissolve and allow integration in settings beyond the socio-cultural context of comics.

6. Conclusion

This analysis has provided four ways in which the visual-graphic and visual-verbal modalities connect. Inherent relations incorporate text and image as part of the others’ structures. Emergent relations interface text and image directly. Adjoined relations integrate text and image without a direct physical interface. Finally, Independent relations keep text and image fully separate. These relationships facilitate differing levels of multimodal integration through the degree to which they unify (or do not unify) the disparate domains into a single expression.

While the tropes of word balloons and thought bubbles have typically been thought of as simple graphic morphemes, they have been shown as surface types illustrating a deeper schema linking images with text (and possibly other images). These Emergent relationships of connectedness facilitate the tight integration of the contents of a Framing component with aspects of a Root representation. This abstract schema for integrating domains becomes manifest using semantic primitives that specify the functional of characteristic of the relationship between the Root and Carrier. Through the mixing of Gestalt principles of perceptual grouping with semantic integration of meaning, text and image are joined into multimodal composite signals. These expressions form bundled cohesive units on par with the mixing of other domains, such as speech and gesture. However, unlike in speech/gesture interactions, text and image do not have a clear dominant and auxiliary relationship. Rather, they feature variability in their connections,

¹ It may be possible to argue that Independent interactions occur when a person draws something and then talks about it. Though it may feature semantic indices in the speech, there is no direct physical connection within either modality to make that integration. Instead, in all likelihood, the manual modality may be employed as an index, where pointing gestures complete the interface.

hinting that human expression does not simply use one dominant modality (verbal) that is supplemented by auxiliary systems, but can use complex multimodal interactions that weigh the contributions of each domain.

7. Acknowledgements

Ray Jackendoff is thanked for his feedback on early drafts of this paper. Funding for this paper was made possible by the Tufts Center for Cognitive Studies.

8. Graphic References

All images are copyright © 2012 Neil Cohn, except those cited throughout the text. Cited images are copyright their respective owners, and are used purely for analytical, critical and scholarly purposes.

Riffenburg, Charles III. *Dumbstruck*. www.grabbagcomics.com. First posted 06/12/07
Schulz, Charles M. 2004. *The Complete Peanuts: 1950-1952*. Edited by G. Groth. Seattle, WA: Fantagraphics Books.

9. References

- Abbott, Lawrence L. 1986. Comic Art: Characteristics and Potentialities of a Narrative Medium. *Journal of Popular Culture* 19, 155-76.
- Arnheim, Rudolf. 1974. *Art and visual perception: A psychology of the creative eye*, Revised edn. Berkeley: University of California Press.
- Bongco, Mila. 2000. *Reading Comics: Language, Culture, and the Concept of the Superhero in Comic Books*. New York, NY: Garland Publishing Inc.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Clark, Herbert H. 1996. *Using Language*. Cambridge, UK: Cambridge University Press.
- Cohn, Neil. 2003. *Early Writings on Visual Language*. Carlsbad, CA: Emaki Productions.
- Cohn, Neil. 2007. A Visual Lexicon [January 2007]. *Public Journal of Semiotics* 1, 53-84.
- Dean, Michael. 2000. *The Ninth Art: Traversing the Cultural Space of the American Comic Book*, University of Wisconsin-Milwaukee dissertation.
- Federmeier, Kara D. & Marta Kutas. 2001. Meaning and modality: influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition* 27, 202-24.
- Forceville, Charles, Tony Veale & Kurt Feyaerts. 2010. Balloonics: The Visuals of Balloons in Comics. In Joyce Goggin & Dan Hassler-Forest (eds.), *The Rise and Reason of Comics and Graphic Literature: Critical Essays on the Form*. Jefferson: McFarland & Company, Inc.
- Ganis, Giorgio, Marta Kutas & Martin I. Sereno. 1996. The search for "common sense": An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience* 8, 89-106.
- Hagan, Susan M. 2007. Visual/Verbal collaboration in print: Complementary differences, necessary ties, and an untapped rhetorical opportunity. *Written Communication* 24, 49-83.

- Hauser, Marc D, Noam Chomsky & W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298, 1569-79.
- Horn, Robert. 1998. *Visual Language: Global Communications for the 21st Century*. Bainbridge Island, WA: MacroVU, Inc.
- Jackendoff, Ray. 1994. *Patterns in the Mind: Language and Human Nature*. New York: Basic Books.
- Magnussen, Anne. 2000. The Semiotics of C.S. Peirce as a Theoretical Framework for the Understanding of Comics. In Anne Magnussen & Hans-Christian Christiansen (eds.), *Comics and Culture: Analytical and Theoretical Approaches to Comics*, 193-207. Copenhagen: Museum of Tusculanum Press.
- McCloud, Scott. 1993. *Understanding Comics: The Invisible Art*. New York, NY: Harper Collins.
- McNeill, David. 1992. *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.
- Mitchell, W.J.T. 1986. *Iconology: Image, Text, Ideology*. Chicago, IL: University of Chicago Press.
- Murray, Chris. 1997. Reading Comics: Narrative and Visual Discourse in Comic Art. Paper presented at Scottish Word and Image Group, University of Dundee.
- Natsume, Fusanosuke. 1997. *Manga wa naze omoshiroi no ka [Why are Manga Fascinating?: Their visual idioms and grammar]*. Tokyo, Japan: NHK Library.
- Palmer, Stephen. 1992. Common region: a new principle of perceptual grouping. *Cognitive Psychology* 24, 436-47.
- Palmer, Stephen & Irvin Rock. 1994. Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin & Review* 1, 29-55.
- Piekos, Nate. 2009. Comics Grammar and Traditions. *Blambot: Comic Fonts and Lettering* (<http://www.blambot.com>). (accessed 12 Nov 2009)
- Pohl, Mary E. D., Kevin O. Pope & Christopher von Nagy. 2002. Olmec Origins of Mesoamerican Writing. *Science* 298, 1984-7.
- Premack, D. & G. Woodruff. 1978. Does the chimpanzee have a 'theory of mind'? *Behavioral and Brain Science* 4, 515-26.
- Saraceni, Mario. 2000. *Language Beyond Language: Comics as Verbo-Visual Texts*. University of Nottingham dissertation.
- Wellman, Henry M., et al. 2002. Thought-bubbles help children with autism acquire an alternative to a theory of mind. *Autism* 6, 343-63.
- Wellman, Henry M., Michelle Hollander & Carolyn A Schult. 1996. Young Children's Understanding of Thought Bubbles and Thoughts. *Child Development* 67, 768-88.
- Wertheimer, Max. 1923. Untersuchungen zur Lehre von der Gestalt (Laws of Organization in Perceptual Forms). *Psychol. Forsch.* 4, 301-50.
- Yannicopoulou, Angela. 2004. Visual Aspects of Written Texts: Preschoolers View Comics. *L1-Educational Studies in Language and Literature* 4, 169-81.