



Klomberg, Bien & Hacimusaoğlu, Irmak & Lichtenberg, Lenneke Doris & Schilperoord, Joost & Cohn, Neil. 2023. Continuity, Co-reference, and Inference in Visual Sequencing. *Glossa: a journal of general linguistics* 8(1). pp. 1–43. DOI: <https://doi.org/10.16995/glossa.9982>



Open Library of Humanities

## Continuity, Co-reference, and Inference in Visual Sequencing

**Bien Klomberg**, Tilburg University, Tilburg School of Humanities and Digital Sciences, Department of Communication and Cognition, Tilburg, The Netherlands, [S.A.M.Klomberg@tilburguniversity.edu](mailto:S.A.M.Klomberg@tilburguniversity.edu)

**Irmak Hacimusaoğlu**, Tilburg University, Tilburg School of Humanities and Digital Sciences, Department of Communication and Cognition, Tilburg, The Netherlands, [I.Hacimusaoglu@tilburguniversity.edu](mailto:I.Hacimusaoglu@tilburguniversity.edu)

**Lenneke Doris Lichtenberg**, Tilburg University, Tilburg School of Humanities and Digital Sciences, Department of Communication and Cognition, Tilburg, The Netherlands, [L.d.Lichtenberg@tilburguniversity.edu](mailto:L.d.Lichtenberg@tilburguniversity.edu)

**Joost Schilperoord**, Tilburg University, Tilburg School of Humanities and Digital Sciences, Department of Communication and Cognition, Tilburg, The Netherlands, [J.Schilperoord@tilburguniversity.edu](mailto:J.Schilperoord@tilburguniversity.edu)

**Neil Cohn**, Tilburg University, Tilburg School of Humanities and Digital Sciences, Department of Communication and Cognition, Tilburg, The Netherlands, [neilcohn@visuallanguagelab.com](mailto:neilcohn@visuallanguagelab.com)

To understand narratives, comprehenders need to establish to what extent various expressions refer to the same entity, known as co-reference. Research in linguistics has long recognized this fundamental aspect of meaning-making, and perhaps as such, an increasing amount of works have attempted to extend this knowledge to other narrative expressions, such as visual storytelling. However, these applications of semantic and discourse models to visual narratives do not sufficiently account for how different visual signals may be reconciled into a common entity. This core question remains unanswered even in theories characterizing meaning relations for visual sequences in comics specifically. Therefore, this paper proposes a formal theory of visual co-reference that accounts for these issues. We present a model of constraints that describes visual co-reference based on varying correspondences between form (graphic input) and meaning (conceptual structure), including subsequent inferences about temporal and spatial relations across panels and more complex structures of visual discontinuity. This theory is in line with both corpus and psychological findings of visual narratives, and thus presents a robust framework for analyses of visual co-reference.

*Glossa: a journal of general linguistics* is a peer-reviewed open access journal published by the Open Library of Humanities. © 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

**OPEN ACCESS**



## 1 Introduction

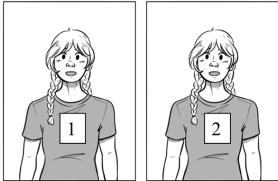


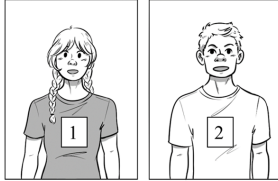
Co-reference has long been recognized as a fundamental aspect of language, pertaining to how comprehenders reconcile expressions referring to the same identity (Chomsky 1980; Gordon & Hendrick 1997; Graesser et al. 1997; Sanders & Gernsbacher 2004). In spoken languages, different conventionalized word forms connect to common identities, despite distinct phonological structures (e.g. proper names vs. pronouns). In sign languages, co-reference can be achieved through signers denoting a particular space in front of them to a particular referent (e.g. person A is to their left, person B is to their right), and then highlighting that space when referring to the respective referent (Frederiksen & Mayberry 2022). While the affordances of these modalities rely on distinct strategies, both achieve the goal of resolving common identities across disparate units (i.e. co-reference). This essential process also occurs in visual storytelling, which has been described using linguistic models. However, while such models may address co-reference in visual sequencing, they typically underestimate the affordances of graphic input, and do not explain why people without exposure to visual narratives may fail to construe co-reference (Cohn 2020). Thus, the current paper uses linguistic formalism to describe the fundamental base that comprehenders seemingly must learn to achieve co-reference in visual language.

Questions of how different modalities achieve similar conceptual processes have been increasingly highlighted within the field of linguistics. For example, approaches like Super Linguistics aim to unify linguistic theories across modalities using formal semantics (Patel-Grosz et al., In press.). Likewise, Visual Language Theory (VLT, see Cohn (2020)) posits that graphic systems draw upon the same structures and cognitive resources as language, including constructions that may need to be learned by comprehenders. This work includes findings that similar neural responses are involved in co-referential processing in visual narratives and language (Coopmans & Cohn 2022), suggesting consistency between the visual and verbal modalities. In line with these approaches and findings, the current work uses established formal linguistic methods to describe co-reference in visual narratives, providing insight into how this conceptual process is achieved within a broader conceptualization of a multimodal language faculty (Cohn & Schilperoord 2022).

Consider **Figure 1**,<sup>1</sup> which compares the affordances of visual and verbal expressions. **Figure 1a-c**'s graphics show two figures across panels that will likely be inferred to be the same person, while **Figure 1d**'s graphics shows figures likely be inferred as different people. In spoken language, these inferences can be achieved by correspondences between an established referent ("Suzy") and other conventionalized expressions (a repetition of the proper noun "Suzy" in **1a** or the pronoun "she" in **1b** and **1c**). Incoherence between the referent and another expression (e.g. between the gender of the pronoun and "Suzy" in **1d**) would lead to disjoint co-reference.

---

<sup>1</sup> This paper presents self-created examples alongside published examples, as both constitute valid productions of visual language. The theory was built upon both types of examples, but self-created panels can show only the variables under discussion without requiring too much context.

a)	
Graphic	
Phonological/syntax	Suzy <sub>1</sub> stood there. Suzy <sub>2</sub> stood there.
Descriptive	[Event BE([Object SUZY] <sub>1</sub> , AT [Place THERE])] [Event BE([Object SUZY] <sub>2</sub> , AT [Place THERE])]
Referential	1                      2 = 1
b)	
Graphic	
Phonological/syntax	Suzy <sub>1</sub> stood there. She <sub>2</sub> frowned.
Descriptive	[Event BE([Object SUZY] <sub>1</sub> , AT [Place THERE])] [Event FROWN([Object SHE] <sub>2</sub> )]
Referential	1                      2 = 1
c)	
Graphic	
Phonological/syntax	Suzy <sub>1</sub> stood there. She <sub>2</sub> turned.
Descriptive	[Event BE([Object SUZY] <sub>1</sub> , AT [Place THERE])] [Event TURN([Object SHE] <sub>2</sub> )]
Referential	1                      2 = 1
d)	
Graphic	
Phonological/syntax	Suzy <sub>1</sub> stood there. He <sub>2</sub> smiled.
Descriptive	[Event STAND([Object SUZY] <sub>1</sub> , AT [Place THERE])] [Event SMILE([Object HE] <sub>2</sub> )]
Referential	1                      2

**Figure 1:** Graphic scenes represented in panel pairs along with similar sentence examples, descriptions of meaning, and referential correspondences. Images © Bien Klomberg.

In visual languages, correspondences must be made based on the graphic input rather than such specific forms (Stainbrook 2016). Visually, the panels in **Figure 1a** use identical lines, supporting these are two drawings of the same object. In **Figure 1b**, slight graphic differences in the figure's face emerge but with enough overall similarity to likewise support inferences of a same object. This inference also holds for **Figure 1c**, despite greater graphic differences due to the woman being shown from the side. A large number of graphic differences also appears in **Figure 1d**, to the extent that we may assume no co-reference between figures. These inferences highlight that any model of visual co-reference needs to explain a large variety of graphic information across panels, that can range from no differences between lines (**Figure 1a**), to slight differences (**Figure 1b**), to greater differences, which may motivate both inferences of a common identity (**Figure 1c**) or not (**Figure 1d**).

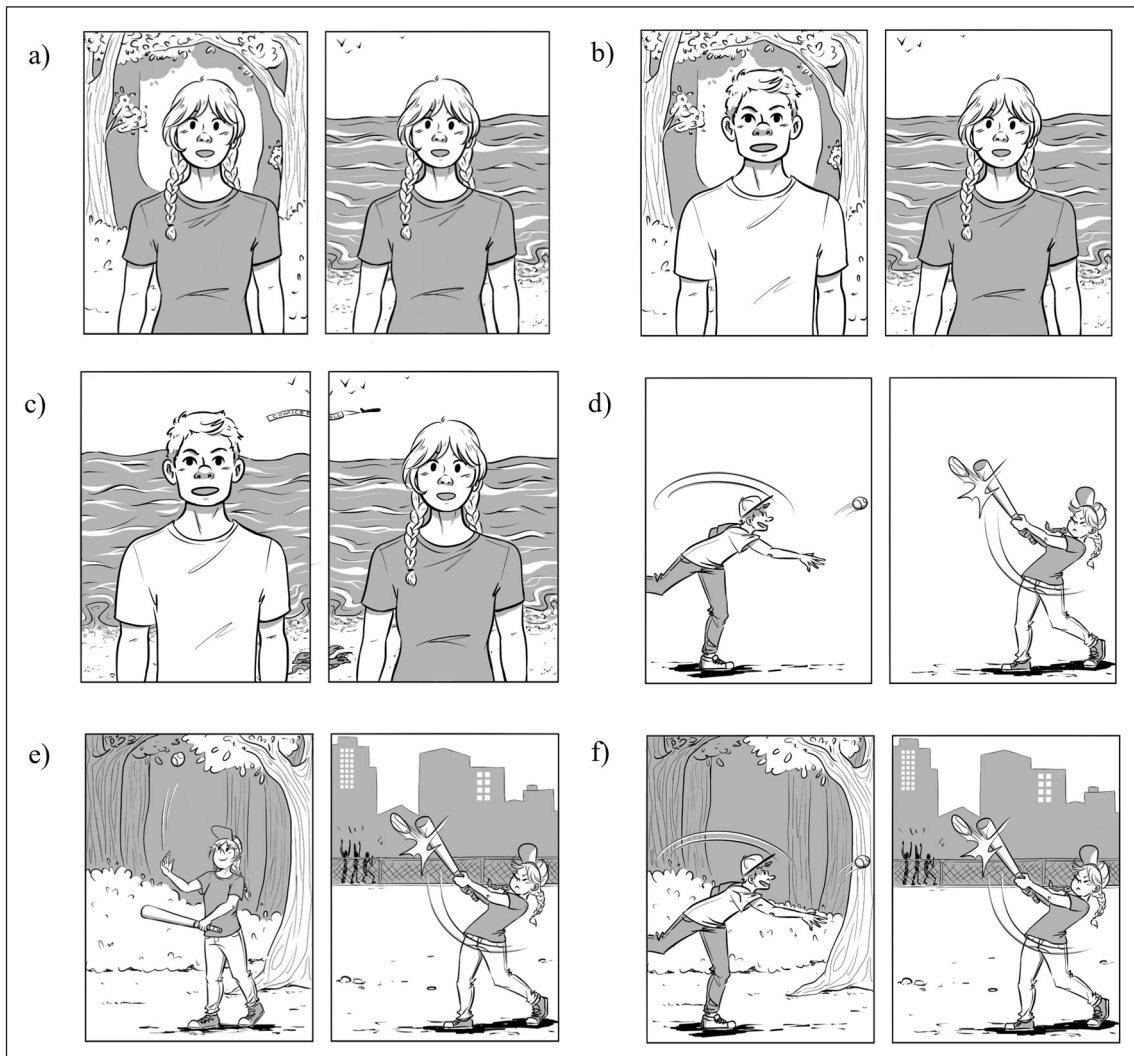
To our knowledge, no model yet formally accounts for this range of graphic comparisons, instead appealing only to general notions of “perception” (Bateman & Wildfeuer 2014; Gavalier & Beavers 2018) or “pragmatics” (Abusch 2012; Maier & Bimpikou 2019; Schlöder & Altshuler 2023). Nevertheless, such co-reference is not recognized by people without exposure to visual narratives; for instance, they may understand a four-panel sequence showing a single person's beard getting longer instead as showing four different but related people, i.e. “brothers” (Núñez 2013, as cited in Cohn 2020). While graphic similarity across images can evidently be perceived, similarity alone appears insufficient to support a relation of co-reference. Moreover, co-referential inferences have been shown to develop in children across a developmental trajectory, with children only recognizing this continuity between ages 4 and 6 years old (Cohn 2020). Just what do people learn (or not learn) that enables the comprehension of co-reference within a visual sequence?

In addition, as many visual narratives are multimodal, comprehenders may also need to balance co-reference between visual and verbal languages. While some discourse-oriented approaches indeed consider such multimodality (Bateman & Wildfeuer 2014; Tseng & Bateman 2018), this does not mitigate the need to account for the ways visual sequences themselves negotiate issues of co-reference, particularly if such a multimodal account is to account for co-reference within and between modalities. Hence, we focus here on wordless visual co-reference, investigating how graphics motivate inferences of common identities.

## 2 Linguistic approaches to visual co-reference

An increasing number of works aims to characterize co-reference in visual narratives using linguistic methods (Abusch 2012; Bateman & Wildfeuer 2014; Braithwaite & Mikkonen 2022; Maier & Bimpikou 2019). Most such theories posit “transitions” between panels based on a key feature of situational change (Gavalier & Beavers 2018; McCloud 1993; Saraceni 2016). Consider **Figure 2**, depicting more complex panel relationships. We already discussed the repetition of

the same character (**Figure 1a**) and changes across characters (**Figure 1d**), but images can also change backgrounds while they maintain (**Figure 2a**) or change characters (**Figure 2b**). **Figure 2c** similarly changes characters, but the slightly different backgrounds here form one continuous scene instead of distinct places. While these three panel pairs (**Figures 2a–2c**) present relatively static moments, the next three sets (**Figures 2d–2f**) show continuity in action. Such events can likewise be shown across different characters (**Figure 2d**), same characters but different backgrounds (**Figure 2e**), or different characters and backgrounds (**Figure 2f**).



**Figure 2:** Panel pairs with incremental discontinuity across characters and locations, shown by shifts between a) locations, b) characters and locations, c) characters, with minimal differences across locations, d) consecutive actions across different characters, e) consecutive actions across different locations, and f) consecutive actions across different characters and locations. Images © Bien Klomberg.

Most panel transitions are proposed to be both binary (they either do or do not occur) and exclusive (only one dimension changes at a time). These theories began most prominently with McCloud’s panel transitions between consecutive events, causes and effects, events in the same scene or semantic field, events across time and space, different facets of a same place, idea, or mood, and last, events with no logical relation. However, in many of the examples in **Figure 2**, multiple dimensions change at once, and incremental change is also prevalent. Corpus studies have also validated that comics use incremental and non-exclusive shifts between panels (Cohn 2023). For example, **Figure 2b** and **2f** show change in both characters and spatial location, and **Figure 2c** presents only minimal differences between scenes. Moreover, **Figure 2d-f** include *partial* change, as the ball is likely inferred to be the same across panels. McCloud (1993) posits that such transitions differ based on the level of “closure” (i.e., inference) the reader must resolve across panels, suggesting a ranking of reader involvement. Relations of causes and effects (such as **Figure 2d**) would require little inference across panels, while transitions across common scenes or semantic fields (e.g. two people seemingly in the same space as in **Figure 1d**) require more. However, when various transitions coincide (e.g. **Figure 2f**), categorizations become limited.

More recent works describe discourse/coherence relations to connect events across panels (Bateman & Wildfeuer 2014; Maier & Bimpikou 2019; Saraceni 2016; Schlöder & Altshuler 2023; Stainbrook 2016). Based on more formal descriptions, such relations support a variety of interpretations gained from visual narratives, including aspects of continuity and co-reference. While these approaches can apply multiple relations and recognize ambiguity (Schlöder & Altshuler 2023), analyses appear concerned with the *conceptual* interpretations of images (the Descriptive tier from **Figure 1**) rather than how that meaning is motivated by the graphic modality. For instance, an analysis of a temporal relation between panels may already presuppose visual co-reference between agents. To construe temporality, we should *first* establish agents to be identical; otherwise, the actions could be performed simultaneously, if we understand the agents to be different people, as in findings from people who do not recognize co-reference in visual sequences. Due to this focus, the question remains how such models would treat the issue of failed co-reference across panels: what do comprehenders lack when they cannot construe such semantic relations?

Indeed most prior models seem to gloss over how graphic cues allow recognition of visual co-reference as a matter of pragmatics or perception. Some describe visual co-reference to rely on “pragmatic enrichment” (Abusch 2012), meaning that comprehenders fill in information not overtly present in the narrative, or handwave it as “purely” pragmatic, with co-reference or disjoint reference being equally viable options (Abusch 2012; Maier & Bimpikou 2019; Schlöder & Altshuler 2023). Others attribute the process of co-reference to models of (natural) perception, such as Gestalt principles (Bateman & Wildfeuer 2014; Gavalier & Beavers 2018). The recognition of drawn objects, and presumably the recognition of sameness of objects through visual

co-reference is, to some extent, thought to be universally accessible as a product of (linguistic) meaning making and/or vision, and thus is not needed to be directly theorized.

However, assumptions of universal transparency contradict empirical findings that imply visual narrative understanding requires exposure and learning. Adults from communities not familiar with (Western) visual narratives may not construe the continuity of characters or events across images (Cohn 2020), while recognition of co-reference across panels appears to emerge in development between 4 and 6 years old (for review see Cohn (2020)). In such cases of lack of proficiency, panels are construed to show multiple separate characters rather than multiple instances of the same character, and/or each picture is treated as its own narrative rather than a connected sequence of events. Since these comprehenders do not seem to have deficits in perception and pragmatics (they function in other communicative settings, e.g. interviews and storytelling), this suggests that understanding visual co-reference requires some level of familiarity with the conventions of that visual language, rather than being construed through properties of perception or pragmatics.

These findings raise several questions: What do comprehenders of visual narrative sequences need to learn in order to understand visual co-reference? How can we formalize dimensions of physical change between visual representations while also acknowledging their continuity (including incremental and/or non-exclusive changes)? How can we account for inferences that arise between juxtaposed units but are not overtly represented in the visuals themselves (as typically reflected in coherence relations)?

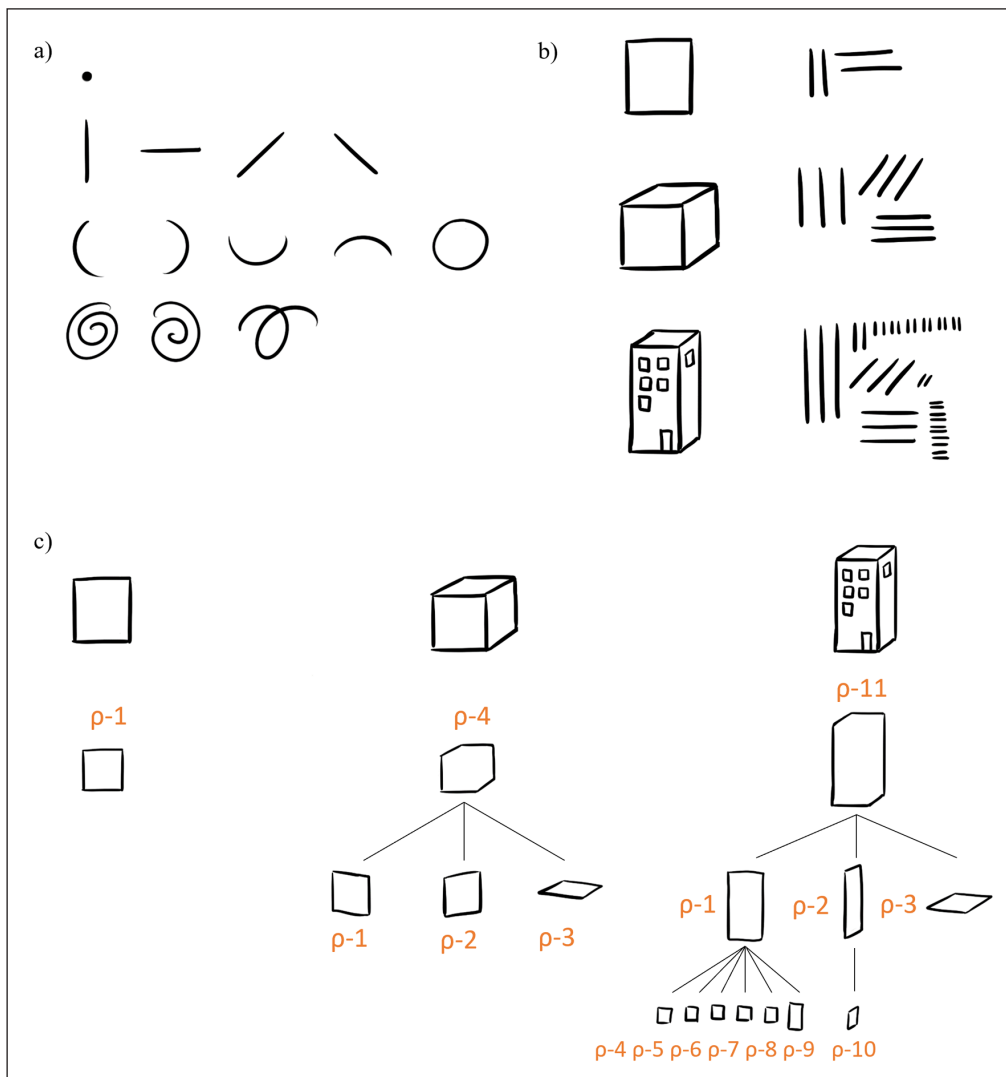
To address these questions, we present a formalized theory of constraints underlying visual (no) co-reference across and within panels, along with subsequent inferences arising between juxtaposed images. Our approach situates these issues as a relationship between form (here, graphics) and meaning, modeled within Jackendoff's Parallel Architecture (Jackendoff 2002) which has also been applied to visual narratives (Cohn 2020). Thus, prior to formalizing our constraints on visual co-reference, we first provide an overview of these two relevant structures of form and meaning: graphological structure and conceptual structure.

### **3 Structures of the model**

#### **3.1 Graphological structure**

If we are to characterize a relationship between form and meaning, we first need to characterize the *form* of pictorial representations (Braithwaite & Mikkonen 2022). We here present a provisional theory of graphological structure, that organizes the visual-graphic aspects of pictorial representations (Cohn 2012), analogous to the phonological structure that organizes speech sounds. We will expand on this theory in later publications but sketch the basic principles to the extent that it can inform our subsequent formalization. At the most basic level, pictorial

representations are composed of lines and shapes in different orientations: dots, lines, curves, and spirals (**Figure 3a**). **Figure 3b** shows how such simple lines can compose various shapes of increasing complexity. Physically produced lines can be considered as “marks”, but they correspond to “graphemes”, which are cognitive “picture primitives” (Willats 1997). This relationship would be similar to that between the uttered speech sounds and the phonemes that represent their corresponding cognitive primitives.



**Figure 3:** Illustrations of a) basic lines and shapes of drawing, b) various shapes and their composite lines, and c) hierarchical regions and their composite regions. Images © Bien Klomberg.



Combinations of lines comprise “regions” (Willats 1997; 2005), which are enclosed visual groupings, whether explicit or inferred, therefore functioning as an abstracted visual variable. We indicate this variable of region with the Greek letter “ $\rho$ ” (rho), consistent with the  $r$  of region (by analogy to the  $\sigma$  used to represent syllables as the primary grouping structure in phonology). Regions can be hierarchic in organization, and in turn may correspond to hierarchies of conceptual volumes that comprise the spatial understanding of objects (e.g., Biederman 1987; Marr 2010) or scenes (Le-Hoa Vö 2021). They are motivated bottom-up by graphics alone, but, like phonological structure, can also be influenced top-down by correspondences to meaning and when situated in context.

We emphasize here that our graphological structure is purely concerned with graphic information (like how phonological structure is purely concerned with sounds) and regions, nor their cognitive graphemes, represent their meaning. Rather, these lines and regions maintain correspondences to conceptual structures which reflect their meaning (Willats 1997).

Let us illustrate this distinction with **Figure 3c**, depicting regional hierarchies for the three images. The first graphic representation forms its own enclosed visual grouping ( $\rho$ -1), as its lines appear to form one “unit” together. The second representation is more complex, as it divides into three parallelograms ( $\rho$ -1,  $\rho$ -2,  $\rho$ -3), each one *mapping* to what would *conceptually* be interpreted as one face of a box (Willats 1997; 2005). Together these graphic lines compose a broader shape marked by the outer contour of the representation ( $\rho$ -4), which then *corresponds* to the encoded concept of a 3D cubic volume (a box). Graphically, both the higher-level contour and lower-level component parts can be identified as regions, just with different internal complexity. Conceptually, these marks correspond to a box. Similarly, the third graphic representation in **Figure 3c** also comprises three regions ( $\rho$ -1,  $\rho$ -2,  $\rho$ -3) but with more internal structure, as  $\rho$ -1 and  $\rho$ -2 encompass other regions ( $\rho$ -4 to  $\rho$ -10). These latter regions map to conceptualizations of “windows” and a “door”, with  $\rho$ -1 and  $\rho$ -2 mapping to the concept of a “wall”. The larger regions along with their internal structure again connect to create the maximal graphic representation reflected by its outer contour, all of which corresponds to the concept of a building.

Such conceptual mappings between form and meaning appear deeply entrenched for comprehenders, making it challenging to separate interpreted meaning from purely graphic input. While we emphasize form-meaning correspondence throughout the paper, for convenience we will at times describe graphic regions (e.g. **Figure 3c**’s regions) as the conceptual objects that those regions map to (e.g., a box and building), to analyze whether those marks show identical or distinct concepts.

### 3.2 Conceptual structure

We here draw on cognitive theories of language and visual narratives (Cohn 2013; Jackendoff 2002), specifically Jackendoff's (1983; 2010) Conceptual Semantics. This approach posits a cognitive conceptual structure organized around primitives of basic ontological categories, which combine using algebraic predicate structures. At the highest level, Conceptual Semantics posits three superordinate ontological categories, which include basic-level conceptual primitives (Jackendoff 2010): Material (Thing, Group, Substance, Aggregate), Space (Place, Path), and Situation (Event, State). These categories are consistent with most cognitive theories of semantics (Gruber 1965; Mandler 2004; Talmy 2003). In our application to visual information, these superordinate categories will generally refer to visual characters or objects, backgrounds, and event information, which we analyze separately to avoid issues of exclusivity. Concepts of Property and Amount also modify or constrain Materials and/or Spaces.

Co-reference involves a particular predicate relation between token or type identity (Jackendoff 1983; 2010). We will use here the notion of *conceptual* token and types, where tokens refer to two instances of a category, while type refers to the category itself. For instance, two entities that are token-identical are considered identical representations, sharing complete co-referentiality (e.g., the same person or object across panels). Type-identical entities are regarded as two separate instances that belong to a same category, maintaining a certain connectedness but not referring to the same object (e.g., two different people, both still 'human', or a knife and a fork as two examples of 'cutlery'). When entities are distinct items from separate categories, these are seen as type-different (e.g., a person and a dog, or a knife and a pillow). Thus, co-reference arises as an inference of continuity between concepts, which here correspond to graphic regions.

## 4 Continuity constraints

Having established a basis for graphic and conceptual structures, we here present constraints guiding the inferences of (no) co-reference across units. Overall, we argue that graphic regions correspond to conceptual structures in the understanding of pictures, and a range of probabilistic constraints then specify the connection of both graphological structures and meaning across images to yield an inference of (no) co-reference. These constraints thus define criteria for when images likely yield co-reference vs. no co-reference, and include top-down processes (e.g. predicting certain information in the second panel based on the first) as well as bottom-up processes. We outline constraints for the continuity of the three superordinate categories of Materials, Space, and Situation, with additional constraints related to the (dis)appearance of visual information. Moreover, we present subsequent inferences that

may follow from these ‘base’ continuity constraints, such as inferences of time, motion, and meaningful discontinuity.

#### 4.1 Material Continuity (MC)

The most straightforward aspect of co-reference in visual sequencing is the continuity made between objects across images, i.e., the understanding that entities in one panel have the same identity as those in another panel. Within conceptual structure, the superordinate category of Materials describes such identifiable entities, with subclasses of Things, Groups, Substances, and Aggregates (Jackendoff 2010), which in visual narratives mostly manifest as graphic depictions of what is conceptually interpreted as characters and/or objects.

We first posit a top-down constraint specifying the expectation that sequential images depict co-referential information. This is specified in MC1. Like all subsequent constraints, we first outline the graphological structure (GS) of the input, which is the graphic information in panels, notated as regions containing elements  $X$  through  $n$  (abstract placeholders for actual graphic input), and distinguished through the superscripts  $a$  and  $b$  as separate panels. Next, we include the input’s conceptual structure (CS), which describes the presence of a character/object in each panel with the abstract  $A$  and  $A'$  (slots that can fill whatever those objects may be, like PERSON), distinguished by superscripts  $\alpha$  and  $\beta$ . Our formalism then provides a probabilistic statement that connects the conditions of graphological structure in the first term with the corresponding conceptual interpretations in the second term. This constraint outlines the process of co-reference for the input and formalizes how both the panels’ graphic information and their corresponding conceptualizations are involved in inferences of co-reference. Subscripts represent co-indexation that bind representations together across structures (see Jackendoff 2002). Namely, conceptual structure indexed by “ $\alpha$ ” is part of the panel indexed by “ $a$ ”, as both are indexed by subscript 1, while conceptual structure indexed by “ $\beta$ ” belongs to the panel indexed by “ $b$ ”, as indexed by subscript 2.

Material Constraint 1 (MC1):

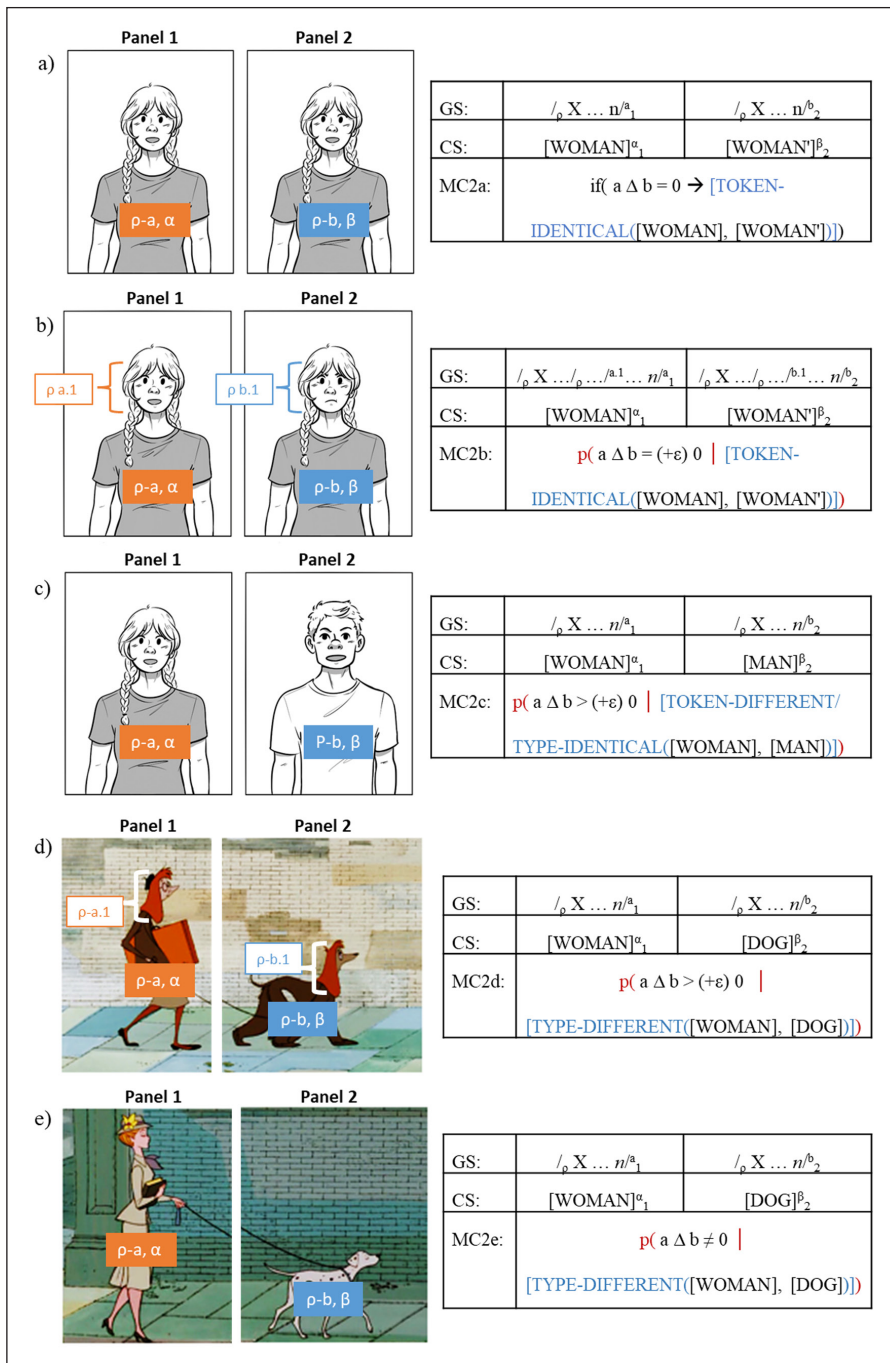
GS:	$/_p X \dots n/_1^a$	$/_p X \dots n/_2^b$
CS:	$[A]_1^\alpha$	$[A]_2^\beta$
MC1:	$p(a, b = [a, b] \mid [\text{TOKEN-IDENTICAL}([\alpha], [\beta])])$	

MC1 as a constraint states the probability ( $p$ ) that when region- $a$  and region- $b$  (panel 1 and 2 respectively) are organized in a juxtaposed sequence (indicated by the square brackets), the inference follows that the corresponding conceptual Material ( $A$  and  $A'$ ) are token-identical.

In other words, presenting panels in a sequence imposes the expectation that the second panel likely shows the same character/object as the first. This top-down structure allows for a forward prediction, consistent with processing models of visual narratives (Cohn 2020). Such a top-down constraint would be built up through familiarity with visual narratives, which tend to repeat characters across panels (Cohn 2020; Saraceni 2016). Nevertheless, this top-down expectancy may not come true, as objects may *not* repeat across panels. This is reflected in our notation of the constraint as a probability, specifying that these inferences have a range of likelihoods. Indeed, violation of this expectation may create costs in processing (Cohn 2020; Loschky et al. 2020), as has been observed in psychological experimentation showing greater updating processes occur with greater situational discontinuity across panels (Cohn 2020). These top-down expectancies can further be modulated by cultural conventions, as distinct patterns of continuity occur for characters, spatial locations and time across American, Asian, and European comics (Klomberg et al. 2022).

Next, we outline bottom-up constraints involving comparisons of graphic regions resulting in inferences about co-reference related to Materials. These comparisons specify a certain amount of change between regions, which ranges from no change at all, to some change, to definite change, and this range comes in different specifications of the broad Material Constraint 2 (MC2). Some variability may occur in these recognized changes, notated by the symbol  $\epsilon$  to account for ambiguity in boundaries (Jackendoff 2010), just as in the phrase “the end of the table”, “the end” is an imprecise boundary that may include a variable amount of table. As outlined in our section on conceptual structure, differences between graphic regions may lead to various inferences of token or type relatedness between the Materials indicated by those graphic signals. One may infer that Materials are identical tokens (the same character), different tokens of a same type (different characters of the same species), or different types altogether (two different species).

Consider the examples in **Figure 4**, starting with **Figure 4a**. Here, the images show two identical sets of lines, likely evoking the idea that these are the same person. This process is captured in MC2a, which describes that *if* the difference (indicated by  $\Delta$ ) between region-a and region-b (which each consist of some unknown  $n$  number of internal regions  $X$ ) is zero, then ( $\rightarrow$ ) the inference follows that the corresponding conceptual Materials ( $A$  and  $A'$ ) are identical. In simpler terms, when there is no difference between the graphics of two regions ( $\rho$ -a and  $\rho$ -b), comprehenders will assume that the characters/objects denoted by those graphics are identical. This constraint therefore assumes the greatest level of continuity (complete co-reference between characters/objects). As applied to **Figure 4a**, we have two regions that conceptually map to a “woman”. When there are no graphic differences between these regions both denoting a woman, then comprehenders assume the two regions show the same woman.



**Figure 4:** Examples with incremental discontinuity for Material information, shown by shifts between a) identical characters, b) highly similar characters, c) different characters, d) different species with some resemblance, and e) different species without resemblance. Figure c and d are slightly adapted screenshots from *101 Dalmatians* © 1996 Walt Disney Pictures; Figure a-c © Bien Klomberg.

Material Constraint 2a (MC2a):

GS:	$/_{\rho} X \dots n^{/a}_1$	$/_{\rho} X \dots n^{/b}_2$
CS:	$[A]^{/a}_1$	$[A']^{/b}_2$
MC2a:	if( $a \Delta b = 0 \rightarrow$ [TOKEN-IDENTICAL( $[a]$ , $[\beta]$ )])	

Next consider **Figure 4b**, where the two images depict largely the same set of lines but with a few differences (highlighted as subregions  $\rho$ -a.1 and  $\rho$ -b.1). We capture this in MC2b: There is a probability ( $p$ ) that when the difference between region-a and region-b is approximately zero (with some variability,  $\epsilon$ ), the corresponding concepts of Materials A and A' are likely interpreted co-referential (token-identical). In short, with minimal physical differences across regions, viewers probably still infer the characters/objects to be the same entity, as illustrated in **Figure 4b** (the graphic differences between two regions both representing a woman are only minimal, meaning these two concepts can be assumed to be the same woman still). Thus, this constraint specifies co-referential continuity is maintained, despite some difference(s) between graphic regions. The variability ( $\epsilon$ ) here would account also for changes in postures, as in **Figure 1c** where the character turned, as graphic regions and their corresponding conceptual interpretations (the woman having braids and a turned-up nose, wearing a short-sleeved shirt, etc.) would constitute enough overlap for co-reference.

Material Constraint 2b (MC2b):

GS:	$/_{\rho} X \dots /_{\rho} \dots /^{/a.1} \dots n^{/a}_1$	$/_{\rho} X \dots /_{\rho} \dots /^{/b.1} \dots n^{/b}_2$
CS:	$[A]^{/a}_1$	$[A']^{/b}_2$
MC2b:	$p(a \Delta b = (+\epsilon) 0 \mid$ [TOKEN-IDENTICAL( $[a]$ , $[\beta]$ )])	

**Figure 4c** depicts distinct differences between lines. Conceptually, readers likely infer these images to show two figures with different faces, hair, and bodies, i.e., separate entities. This is described in MC2c, which states a probability that when the difference between region-a and region-b is more than zero (with some variability), readers likely infer the corresponding concepts of A and A' to be type-identical at most. This captures an increasing level of discontinuity, where relatively more differences become visible between the two regions, and thus we interpret these as distinct instances of a same category type. This constraint could apply to different people (still part of the same category of “humans”), animals, types of furniture,

etc. In **Figure 4c**, both figures have rough correspondence between regions, in that the outer contours of the regions correspond to the concepts of a head, a neck, shoulders, etc. for both panels. However, the evident variations in those regions (e.g. a broader region denoting a thicker neck in panel 2) and in their internal structure (differences in lines denoting different eyes, nose, mouth) constitute enough graphic change for the impression that the figures are distinct tokens.

Material Constraint 2c (MC2c):

GS:	$/\rho X \dots n/\alpha_1$	$/\rho X \dots n/\beta_2$
CS:	$[A]^\alpha_1$	$[A]^\beta_2$
MC2c:	$p(a \Delta b > (+\epsilon) 0 \mid [\text{TOKEN-DIFFERENT/ TYPE-IDENTICAL}([\alpha], [\beta])])$	

For the next two constraints, consider the examples in **Figure 4d** and **4e**. For **4d**, there is sufficient discontinuity across regions to infer these as type-different instances (conceptually, a person and a dog), yet the apparent similarity between various subregions (e.g., the hair on top of the entities' heads, notated as subregions  $\rho$ -a.1 and  $\rho$ -b.1) elicits some relation between regions still (see e.g., Lichtenberg et al. 2022). For **4e**, however, no such resemblance persists between regions, leading to significant differences in graphological structure across regions. We capture this discontinuity (i.e., no co-reference) between regions in constraints that interpret them as type-different instances. The first of these two constraints (MC2d) describes a difference between regions that again is more than zero (with some variability). The second (MC2e) states that the difference in regions is not zero, meaning there is unquestionable change.

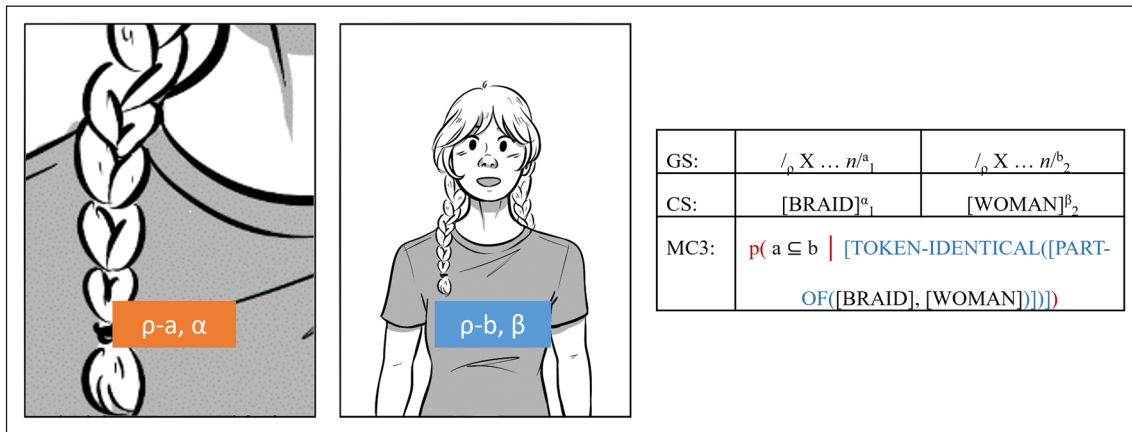
Material Constraint 2d-e (MC2d-e):

GS:	$/\rho X \dots n/\alpha_1$	$/\rho X \dots n/\beta_2$
CS:	$[A]^\alpha_1$	$[A]^\beta_2$
MC2d:	$p(a \Delta b > (+\epsilon) 0 \mid [\text{TYPE-DIFFERENT}([\alpha], [\beta])])$	
MC2e:	$p(a \Delta b \neq 0 \mid [\text{TYPE-DIFFERENT}([\alpha], [\beta])])$	

The distinction between these two versions is that constraint MC2d includes some hint of similarity, e.g. the Materials may share similar contours, colors, or poses, while MC2e has the impression of differences only. Moreover, one may notice that the notation for MC2d's

graphological input is the same as MC2c's, despite different conceptual inferences. Due to the wide range of variability in visual sequencing, minimal graphological similarity across inputs can potentially yield interpretations of both conceptual relations, as in **Figure 4c** and **4d**. It depends on this variability (notated by  $\epsilon$ ) which output will apply.

Finally, consider instances where a representation is not shown in its entirety in a panel but may be preceded or followed by a full view of the representation, leading to the realization that one panel shows part of the whole. This occurs in **Figure 5**, where a close-up of a braid in a panel precedes a panel showing the same braid together with the rest of the woman.



**Figure 5:** Panel pair exemplifying a part-whole relation.

We account for this partitive relationship in MC3: This constraint specifies the probability that when region-a is included in region-b, the two representations are interpreted as identical objects, with the first panel showing part of the larger whole. This formula can be altered depending on whether the panel showing the whole representation comes before ( $\supseteq$ ) or after ( $\subseteq$ ) the partial representation.

Material Constraint 3 (MC3):

GS:	$/_{\rho} X \dots n^{\alpha}_1$	$/_{\rho} X \dots n^{\beta}_2$
CS:	$[A]^{\alpha}_1$	$[A]^{\beta}_2$
MC3:	$p(a \subseteq b \mid [TOKEN-IDENTICAL([PART-OF([A], [B])]))$	

To summarize, these various Material constraints account for a general likelihood of continuity across panels and describe a range of interpretations for the actual input's co-reference based on



images' graphics. Most of this range is captured in MC2's iterations, which gradually move from the most continuous graphic information and complete co-referentiality to increased discontinuity that ultimately leads to no co-reference. Thus, these constraints can be ranked along a scale of (dis)continuity, with MC2a as most continuous and MC2e as most discontinuous. Consequently, when images combine elements with varying levels of continuity, we may assume that any graphic element eliciting discontinuity outweighs the more continuous graphic components in the ultimate impression of the image, as in a preference rule system (Jackendoff 1983; Lerdahl & Jackendoff 1983). As such, we may posit a ranking constraint, which notates the probability that when graphic information evokes multiple constraints (two or more of MC2a-e), the more discontinuous instances are superior to more continuous ones (shown by the ordered sequence within the squared brackets).

Ranking constraint:

$$p(\{MC2a, MC2b, MC2c, MC2d, MC2e\} \mid [MC2e > MC2d > MC2c > MC2b > MC2a])$$

## 4.2 Spatial Continuity (SpC)

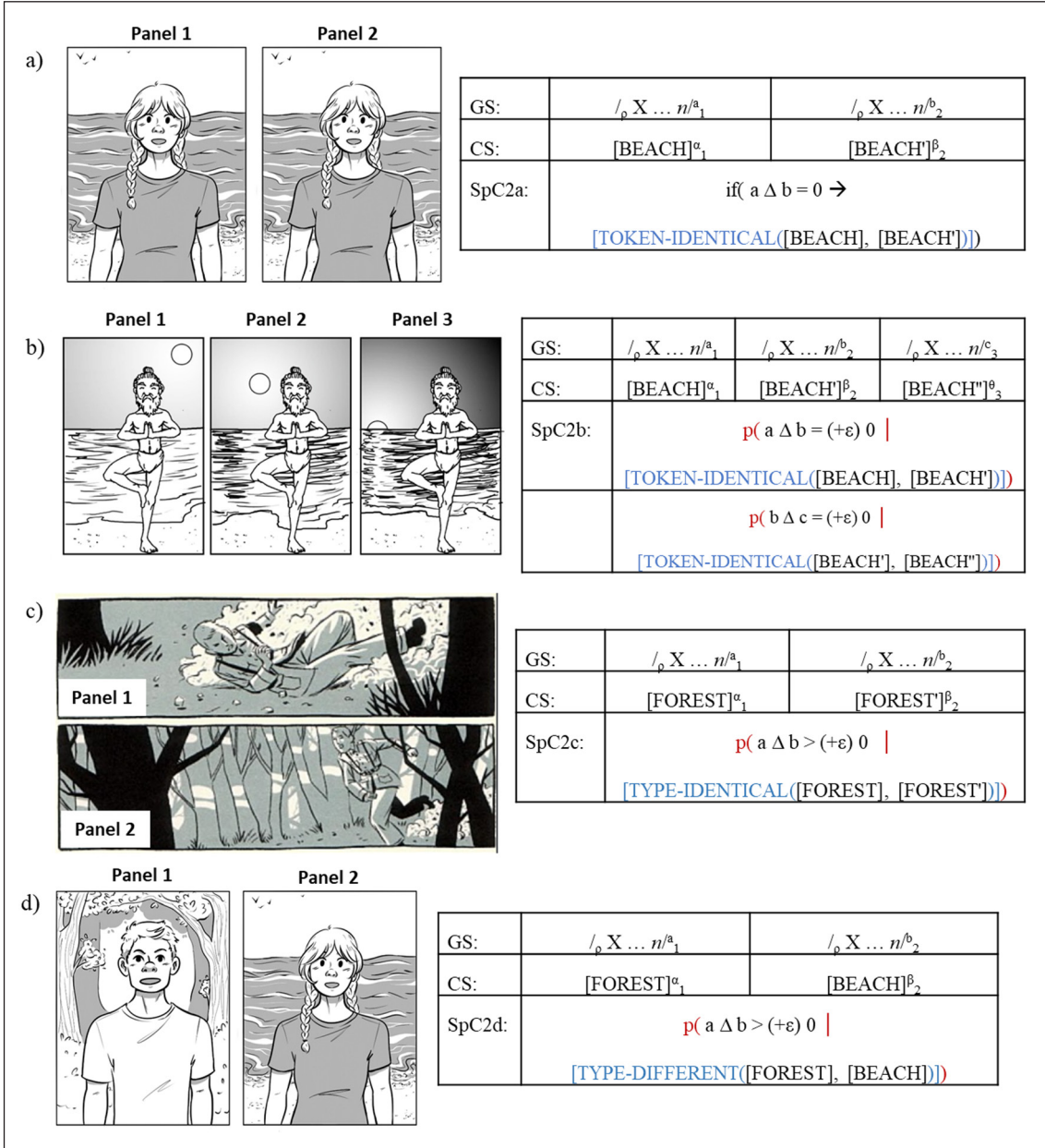
The superordinate category of Space includes Places and Paths (Jackendoff 2010), which in visual narratives often translates to spatial information conveyed through backgrounds in panels. Overall, these spatial constraints mimic those for Materials, as these are also ordered along a scale of continuity. This category includes a top-down constraint (SpC1) specifying the likelihood that sequential images will maintain shared Spaces.

Spatial Constraint 1 (SpC1):

GS:	$/_{\rho} X \dots n^{a_1}$	$/_{\rho} X \dots n^{b_2}$
CS:	$[A]_1^a$	$[A]_2^b$
SpC1:	$p(a, b = [a, b] \mid [TOKEN-IDENTICAL([\alpha], [\beta])])$	

This constraint again states that when region-a and region-b are organized in a sequence, there is the probability that we infer the two Spaces to convey the same location. By specifying a sequential constraint in this fashion, it again allows for a top-down expectancy of continuity between panels. This expectation reflects that visual narratives typically maintain the same spatial location for several panels in a row (Klomberg et al. 2022). Following this top-down likelihood of continuity, we can again posit constraints based on the degree of difference in

bottom-up visual information across regions. **Figure 6** illustrates examples corresponding to these bottom-up constraints.



**Figure 6:** Examples with incremental discontinuity for Spatial information, shown by shifts between a) identical locations, b) highly similar locations, c) related locations, d) distinct locations. Figure c is slightly adapted from the graphic novel *El Faro* © 2004 Paco Roca; Figure b © Neil Cohn; Figure a and d © Bien Klomberg.

Next, we outline the bottom-up continuity constraints (SpC2a-d), which largely follow the same structure as the constraints for Materials. As such we can list them all at once:

Spatial Constraint 2a-d (SpC2a-d):

GS:	$/_p X \dots n/^{a}_1$	$/_p X \dots n/^{b}_2$
CS:	$[A]^{a}_1$	$[A]^{b}_2$
SpC2a:	if( $a \Delta b = 0 \rightarrow [\text{TOKEN-IDENTICAL}([\alpha], [\beta])]$ )	
SpC2b:	$p(a \Delta b = (+\varepsilon) 0 \mid [\text{TOKEN-IDENTICAL}([\alpha], [\beta])])$	
SpC2c:	$p(a \Delta b > (+\varepsilon) 0 \mid [\text{TYPE-IDENTICAL}([\alpha], [\beta])])$	
SpC2d:	$p(a \Delta b > (+\varepsilon) 0 \mid [\text{TYPE-DIFFERENT}([\alpha], [\beta])])$	

SpC2a again expresses the highest level of continuity, namely that when the difference between regions is exactly zero (meaning the regions do not differ at all), then we infer the Spaces corresponding to those regions to be exactly the same. This is depicted by **Figure 6a**, where the background scenery across panels is identical. SpC2b describes some minor difference between regions, that still elicits the impression of a token-identical Space, as in the example in **Figure 6b**. Here, only the waterline and position of the sun change, so the overall sense of continuity across Spaces persists.

Next, SpC2c posits that the difference between regions is now larger than zero and likely leads to the inference that the two corresponding scenes are not the same exactly, but part of a same category instead. We consider Spaces to be type-identical when they exhibit environments that are part of the same category (e.g., both are restaurants, schools, or forests) or are visually related as distinct parts of the same Space (e.g., different parts of the same room). This occurs in **Figure 6c**, where a man runs through the forest and each panel displays another part of the forest.

The final constraint (SpC2d) then describes that relatively large differences between regions probably result in inferring two distinct locations, as shown in **Figure 6d**. This forest and beach appear non-related and are inferred to be separate locations. As in the Material constraints, the same graphological notation applies to SpC2c and SpC2d due to the variability found in visual sequences. Depending on this variability ( $\varepsilon$ ), minimal similarity across inputs can lead to either conceptual interpretation, exemplified in **Figure 6c** and **6d**.

### 4.3 Situational Continuity (SC)

The next set of constraints applies to the superordinate category of Situations, encompassing Events and States (Jackendoff 2010). Similar to the other categories, we first describe a top-down constraint, followed by three continuity constraints fed by bottom-up information.

Situational Constraint 1 (SC1):







GS:	$/_{\rho} X \dots n/^{a}_1$	$/_{\rho} X \dots n/^{b}_2$
CS:	$[A]^{a}_1$	$[A']^{b}_2$
SC1:	$p(a, b = [a, b] \mid \text{TOKEN-DIFFERENT/}$ $\text{TYPE-IDENTICAL}([\text{Situation F}([\alpha]), [\text{Situation F}([\beta])]))$	

This top-down constraint states that when region-a and region-b are ordered one after the other, there is a probability that we infer them as different but related events. Namely, the common prediction for visual sequences is that each panel shows a *different* part of an event, ultimately conveying consecutive moments (Klomberg et al. 2022). This top-down expectation of temporal sequentiality would be the constraint that assumes that sequences of images convey successive moments by default, such as McCloud’s (2000) notion of a “temporal map”. Temporality is not a default for construing visual sequencing (Cohn 2010), and the notion of “time” overall arises as an inference made about successive events. Such a probabilistic constraint as in SC1 can express this general expectancy for successive events, even if it is not a default principle of visual sequencing.

The following two bottom-up constraints make similar comparisons across regions: The difference between region-a and region-b is not zero, with some room for variability. Take **Figure 7a**, where the characters’ postures change across panels, but both express the event of “walking”. In contrast, **Figure 7b** shows a woman performing the action of hitting a ball, while the man may be inferred to be just watching. His relation to her actions is merely inferred, and it is possible that these panels occur at the same “moment” despite being categorically different events.

These differences can be expressed in constraints SC2a and SC2b. Visual differences occur across regions, which for SC2a elicits the inference that the two Situations denoted by those graphic signals are distinct but related to an overarching action (as in **Figure 7a**). This

would also encompass **Figure 2d-f**, where a person threw a ball and the same/a different person hit it. Throwing the ball and swinging the bat would be united as parts of the broader action “batting”. On the other hand, constraint SC2b leads to the inference that the difference in regions elicits separate Situations expressing different events/actions (as in **Figure 7b**).

a)	Panel 1	Panel 2	GS:	$/_p X \dots n^a_1$	$/_p X \dots n^b_2$	
			CS:	$[WOMAN]^{\alpha_1}$	$[WOMAN]^{\beta_2}$	
			SC2a:	$p(a \Delta b \neq (+\epsilon) 0 \mid \text{TOKEN-DIFFERENT/}$ $\text{TYPE-IDENTICAL}([_{\text{Situation}} \text{WALK}([WOMAN])],$ $[_{\text{Situation}} \text{WALK}([WOMAN])]))$		
b)	Panel 1	Panel 2	GS:	$/_p X \dots n^a_1$	$/_p X \dots n^b_2$	
			CS:	$[MAN]^{\alpha_1}$	$[WOMAN]^{\beta_2}$	
			SC2b:	$p(a \Delta b \neq (+\epsilon) 0 \mid \text{TYPE-DIFFERENT}([_{\text{Situation}} \text{BE}([MAN]), [_{\text{Situation}} \text{HIT}([WOMAN, BALL])]))$		
c)	Panel 1	Panel 2	GS:	$/_p X \dots n^a_1$	$/_p X \dots n^b_2$	$/_p X \dots n^c_3$
			CS:	$[MAN]^{\alpha_1}$	$[MAN]^{\beta_2}$	$[MAN]^{\gamma_3}$
			SC2c:	$p(a \Delta b = (+\epsilon) 0 \mid [\text{TOKEN-IDENTICAL}([_{\text{Situation}} \text{MEDITATE}([MAN]), [_{\text{Situation}} \text{MAN}'([MEDITATE])]))$ $p(b \Delta c = (+\epsilon) 0 \mid [\text{TOKEN-IDENTICAL}([_{\text{Situation}} \text{MEDITATE}([MAN']), [_{\text{Situation}} \text{MAN}''([MEDITATE])]))$		

**Figure 7:** Examples with varying levels of (dis)continuity for Situational information, shown by shifts between a) related actions, b) different actions, and c) identical actions. Figure a are slightly adapted screenshots from *101 Dalmatians* © 1996 Walt Disney Pictures; Figure b © Bien Klomberg; Figure c © Neil Cohn.

Situational Constraint 2a-c (SC2a-c):

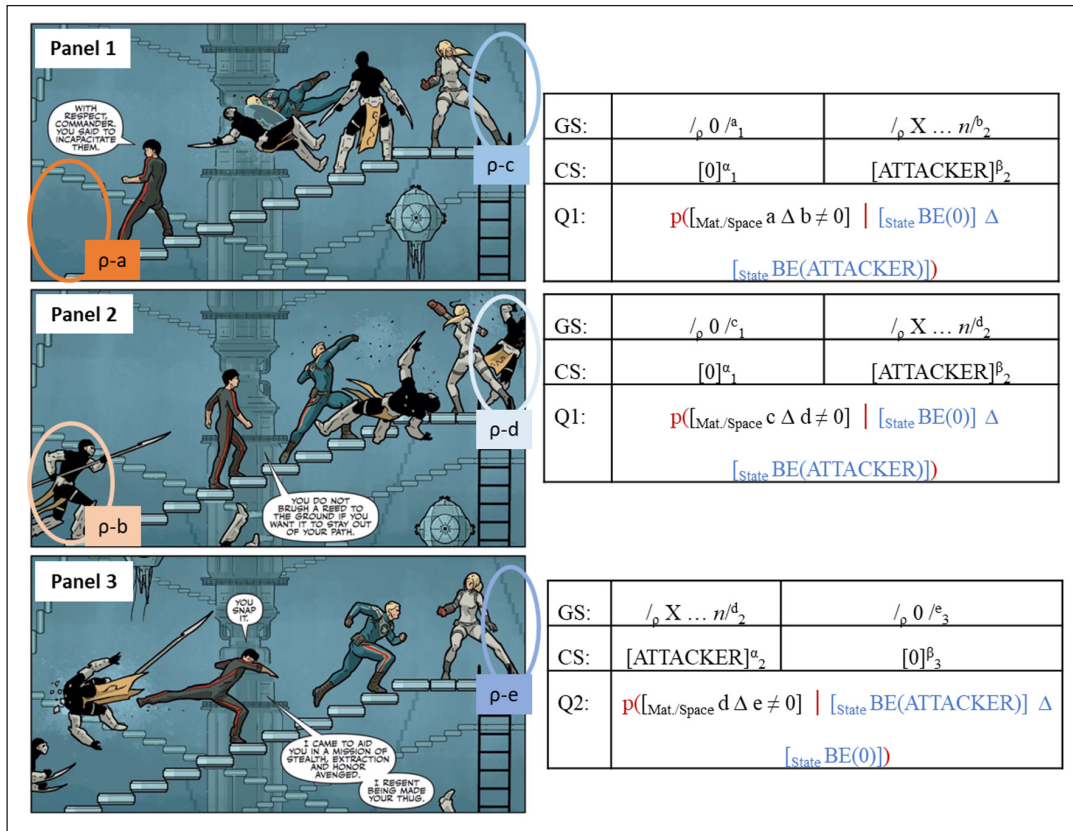
GS:	$/_p X \dots n/^a_1$	$/_p X \dots n/^b_2$
CS:	$[A]^a_1$	$[A]^b_2$
SC2a:	$p(a \Delta b \neq (+\epsilon) 0 \mid \text{TOKEN-DIFFERENT/ TYPE-IDENTICAL}([ \text{Situation F}([\alpha]), [ \text{Situation F}([\beta]) ] ]))$	
SC2b:	$p(a \Delta b \neq (+\epsilon) 0 \mid \text{TYPE-DIFFERENT}([ \text{Situation F}([\alpha]), [ \text{Situation F}([\beta]) ] ]))$	
SC2c:	$p(a \Delta b = (+\epsilon) 0 \mid \text{[TOKEN-IDENTICAL}([ \text{Situation F}([\alpha]), [ \text{Situation F}([\beta]) ] ]))$	

The final bottom-up constraint (SC2c) posits that the difference between regions is approximately zero, meaning that there is no, or only minimal, differences across regions. This leads to the interpretation that the events/states denoted by those regions are the same. For example, in **Figure 7c**, the exact same posture is repeated across panels, suggesting that the figure does not move. Any inferences of time passing are instead motivated by visual differences across the Spatial environment.

#### 4.4 Constraints related to quantities

So far, we have presented constraints that consider the differences in two sets of input (regions) to make an inference about their conceptual co-reference. However, sequences of panels may not always maintain the same number of characters or scenes throughout. Panels may suddenly introduce characters or background scenery or omit these elements from one panel to the next. Indeed, corpus analyses on comics have demonstrated partial changes persisting substantially across visual sequencing, particularly with additions or omissions of which characters are shown in panels (Cohn 2020; Klomberg et al. 2022). When comparing across panels that include a disparate number of Materials and/or Spaces, it is necessary to specify when these elements (dis)appear, which is accounted for in quantity constraints.

Consider **Figure 8**, where a few (dis)appearances occur across panels: From the first to second panel, two new characters are introduced in the second image. Region-a and region-c in the first image can thus be regarded as ‘empty’ regions, where no visual input existed yet. The second image has graphic information showing two characters (region-b and region-d). The change between region-a and region-b and between region-c and region-d shows that characters were added that were not there previously.



**Figure 8:** Example showing the addition and omission of figures across panels. Image is slightly adapted from *Secret Avengers #18* © 2012 Marvel Comics.

First, we address the addition of characters that were not in prior panels. This is expressed in Quantity constraint 1 (Q1).

Quantity constraint 1 (Q1):

GS:	$/_{\rho} 0 /_{\alpha_1}$	$/_{\rho} X \dots n /_{\beta_2}$
CS:	$[0]_{\alpha_1}$	$[A]_{\beta_2}$
Q1:	$p([\text{Mat./Space } a \Delta b \neq 0] \mid [\text{State BE}(\alpha)] \Delta [\text{State BE}(\beta)])$	

This constraint expresses that region-a includes no Material or Spatial information, indicated by 0, and that region-b does have input. When there is such a difference from the absence of information in region-a to visible information in region-b, there is a probability that we infer the sequence goes from a state of there being nothing (0) to a state where there is something (a conceptual element A).

A second Quantity Constraint (Q2) can express the reverse phenomenon, where elements are omitted across panels.

Quantity constraint 2 (Q2):

GS:	$/_{\rho} X \dots n /_1^a$	$/_{\rho} 0 /_2^b$
CS:	$[A]_1^a$	$[0]_2^b$
Q2:	$p([\text{Mat./Space } a \Delta b \neq 0] \mid [\text{State BE}(\alpha)] \Delta [\text{State BE}(\beta)])$	

Q2 states the opposite of Q1: There is a probability that when the visual input in region-a (region-d in **Figure 8**) aligns with no information in region-b (region-e in **Figure 8**), we infer the conceptual Material or Spatial aspect (A) to change into being absent (0). In other words, the visual input present in the previous panel disappeared. As illustrated in **Figure 8**: From the second to the third panel, the character shown by region-d was omitted from the image. Thus, we infer this Material went from a state of being visually depicted to having no input related to them anymore.

## 5 Subsequent inferences

The constraints so far describe the ways that we establish inferences of continuity based on graphic representations across panels. The outcomes of those inferences represent the extent of co-referentiality between elements of three individual categories of information. These impressions of continuity do not operate in isolation and may combine to elicit additional inferences that characterize relations between panels. Furthermore, more complex visual sequences require additional inferencing to be understood, such as sequences that include depictions of “unreal” events, such as imagination or dream sequences. In this section, we describe further construals for how these cases can be understood.









### 5.1 Progression construals (PR)

So far, we considered the continuity of Materials, Spaces, and Situations in isolation, but panels naturally combine these categories as they typically depict characters within a certain location and performing a certain action. Such combinations of co-reference relations may evoke additional inferences. For example, when we recognize that two panels show the same character and that they are involved in related actions, we may interpret that the panels show two consecutive moments. Likewise, one character shown with two related but slightly different backgrounds would seem to have moved from one spot to another.

Consider the examples in **Figure 9**. Here we provide an analysis for each superordinate category, which together support a subsequent inference of how the panels may be related. The first two examples present static scenes across different environments, where the characters remain the same (**Figure 9a**) or change (**9b**). This static impression is due to the postures reflecting identical



Situations. The identical character in **9a** allows for the interpretation that a same figure moved from one place to another, tying the panels together as a movement through locations. **9b**, however, has no such continuity across characters, which leads to the impression of two separate scenes. The other two examples present a contiguous action across panels, due to the Situations now being related actions, with either a same character and environment (**Figure 9c**) or distinct characters and places (**Figure 9d**). Consequently, **Figure 9c** seems to present two consecutive moments. **Figure 9d** also elicits such a temporal shift, but also includes a spatial change, eliciting the sense of a consecutive moment occurring across different locations with different people. The variations within combinations of situational changes may thus lead to different subsequent inferences of panel relations.

a)			b)																						
	<table border="1"> <thead> <tr> <th colspan="2">Panel 1 – 2</th> </tr> </thead> <tbody> <tr> <td>Materials</td> <td>Token-identical (MC1, MC2a)</td> </tr> <tr> <td>Space</td> <td>Type-different (SpC2d)</td> </tr> <tr> <td>Situation</td> <td>Token-identical (SC2c)</td> </tr> <tr> <td>Subsequent inference</td> <td>Spatial progression (PR1), supported by continuity of Materials, identical Situations, and discontinuity of Space.</td> </tr> </tbody> </table>		Panel 1 – 2		Materials	Token-identical (MC1, MC2a)	Space	Type-different (SpC2d)	Situation	Token-identical (SC2c)	Subsequent inference	Spatial progression (PR1), supported by continuity of Materials, identical Situations, and discontinuity of Space.		<table border="1"> <thead> <tr> <th colspan="2">Panel 1 – 2</th> </tr> </thead> <tbody> <tr> <td>Materials</td> <td>Type-identical (MC2c)</td> </tr> <tr> <td>Space</td> <td>Type-different (SpC2d)</td> </tr> <tr> <td>Situation</td> <td>Token-identical (SC2c)</td> </tr> <tr> <td>Subsequent inference</td> <td>Separate situations (ST2), supported by discontinuity of Materials and Space, and identical Situations.</td> </tr> </tbody> </table>		Panel 1 – 2		Materials	Type-identical (MC2c)	Space	Type-different (SpC2d)	Situation	Token-identical (SC2c)	Subsequent inference	Separate situations (ST2), supported by discontinuity of Materials and Space, and identical Situations.
Panel 1 – 2																									
Materials	Token-identical (MC1, MC2a)																								
Space	Type-different (SpC2d)																								
Situation	Token-identical (SC2c)																								
Subsequent inference	Spatial progression (PR1), supported by continuity of Materials, identical Situations, and discontinuity of Space.																								
Panel 1 – 2																									
Materials	Type-identical (MC2c)																								
Space	Type-different (SpC2d)																								
Situation	Token-identical (SC2c)																								
Subsequent inference	Separate situations (ST2), supported by discontinuity of Materials and Space, and identical Situations.																								
c)			d)																						
	<table border="1"> <thead> <tr> <th colspan="2">Panel 1 – 2</th> </tr> </thead> <tbody> <tr> <td>Materials</td> <td>Token-identical (MC1, MC2b)</td> </tr> <tr> <td>Space</td> <td>Token-identical (SpC1, SpC2b)</td> </tr> <tr> <td>Situation</td> <td>Type-identical (SC1, SC2a)</td> </tr> <tr> <td>Subsequent inference</td> <td>Temporal progression (PR1), supported by continuity across all three categories. Also, progression informed by composition (CPR1) for the ball.</td> </tr> </tbody> </table>		Panel 1 – 2		Materials	Token-identical (MC1, MC2b)	Space	Token-identical (SpC1, SpC2b)	Situation	Type-identical (SC1, SC2a)	Subsequent inference	Temporal progression (PR1), supported by continuity across all three categories. Also, progression informed by composition (CPR1) for the ball.		<table border="1"> <thead> <tr> <th colspan="2">Panel 1 – 2</th> </tr> </thead> <tbody> <tr> <td>Materials</td> <td>Type-identical (MC2c)</td> </tr> <tr> <td>Space</td> <td>Type-different (SpC2d)</td> </tr> <tr> <td>Situation</td> <td>Type-identical (SC1, SC2a)</td> </tr> <tr> <td>Subsequent inference</td> <td>Temporal and spatial progression (PR1), supported by discontinuity of Materials and Space and continuity of Situation.</td> </tr> </tbody> </table>		Panel 1 – 2		Materials	Type-identical (MC2c)	Space	Type-different (SpC2d)	Situation	Type-identical (SC1, SC2a)	Subsequent inference	Temporal and spatial progression (PR1), supported by discontinuity of Materials and Space and continuity of Situation.
Panel 1 – 2																									
Materials	Token-identical (MC1, MC2b)																								
Space	Token-identical (SpC1, SpC2b)																								
Situation	Type-identical (SC1, SC2a)																								
Subsequent inference	Temporal progression (PR1), supported by continuity across all three categories. Also, progression informed by composition (CPR1) for the ball.																								
Panel 1 – 2																									
Materials	Type-identical (MC2c)																								
Space	Type-different (SpC2d)																								
Situation	Type-identical (SC1, SC2a)																								
Subsequent inference	Temporal and spatial progression (PR1), supported by discontinuity of Materials and Space and continuity of Situation.																								

**Figure 9:** Analyses of panel pairs with incremental discontinuity across characters and spatial locations, with shifts between a) locations, b) characters and locations, c) consecutive actions, and d) consecutive actions across different characters and locations.

We summarize these types of subsequent inferences as construals of progression across panels (see also Napoli and Leeson (2020)). With *progression* we refer to changes in states that can be characterized by temporal or spatial relations. We first present a general construal that may apply to both time and space. This construal may base itself on the combination of category memberships assigned to the superordinate categories or on differences between regions. Based on either (or sometimes both) of these, the inference arises that the two conceptual items (A and A'), each depicted in a certain Situation or Place (marked by Y in the formula), are connected together by an overarching Situation of progression (the function “GO”, see Jackendoff (1983)).

Progression construal 1 (PR1):

GS:	$/_{\rho} X \dots n^a_1$	$/_{\rho} X \dots n^b_2$
CS:	$[A]^a_1$	$[A]^b_2$
PR1:	$p([_{\text{Mat./ Sit./ Space}} \text{CATEGORY-MEMBERSHIP}(\alpha, \beta)] \mid [_{\text{Situation}} \text{GO}_{\text{Event}}([_{\text{Mat./ Sit.}} X],$ $[_{\text{Path}} \text{FROM}([_{\text{Y}} F(\alpha)), \text{TO}([_{\text{Y}} F(\beta))]))])$ $p([_{\text{Mat./ Sit./ Space}} a \Delta b = (+\epsilon) 0] \mid [_{\text{Situation}} \text{GO}_{\text{Event}}([_{\text{Mat./ Sit.}} X], [_{\text{Path}} \text{FROM}([_{\text{Y}} F(\alpha)),$ $\text{TO}([_{\text{Y}} F(\beta))]))])$	

The Y's here can be filled in by an *event* or *state* when it concerns temporal progressions. In that case, one object goes from one event/state to another, implying a progression in time. The general function F is filled in by the specific event description that applies to each piece of conceptual information (A and A'). For example, in the panel pair in **Figure 9c**, we experience an impression of time passing, as the woman first throws the ball in the air and then hits it. This inference is made possible by complete co-reference across Materials and Space, and overall continuity of Situation. Namely, we see one character who remains in one location, shown in two states that go together as parts of one larger action. This inferred temporal progression would be:

$$[_{\text{Situation}} \text{GO}_{\text{Event}}([_{\text{Situation}} X], [_{\text{Path}} \text{FROM}([_{\text{Event}} \text{THROW}(\text{WOMAN}, \text{BALL}))], \text{TO}([_{\text{Event}} \text{HIT}(\text{WOMAN}, \text{BALL}))])])]$$

The Y may also be filled in by a *place*, in case of spatial progression. This applies to the panel pair in **Figure 9a**, where we recognize one character across distinct locations. The Material

continuity and Spatial discontinuity leads to the inference that one character moved across separate locations, which would result in:

$$[_{\text{Situation}} \text{GO}_{\text{Event}} ([_{\text{Material}} \text{WOMAN}], [_{\text{Path}} \text{FROM} ([_{\text{Place}} \text{IN}(\text{FOREST}))], \text{TO} ([_{\text{Place}} \text{ON}(\text{BEACH}))])])].$$

Inferences of temporal and spatial progression can also coincide, as in **Figure 9d** where Situational continuity combines with Spatial discontinuity. Here, Material and Spatial dimensions lack any co-reference, while the depicted states across panels can be united as parts of a larger action. This continuity allows the interpretation that the narrative switches across two consecutive moments in time that occur in distinct locations, performed by different agents. The actions in **Figure 9b** cannot be united as such, and instead appear identical, which, together with the overall discontinuity in Materials and Space, may elicit the inference that there is no relation across panels.

The category membership for the superordinate categories is not the only element that supports inferences of motion. Consider **Figure 9c** again, where the first image shows a ball high up, while the second image shows it further down. This difference in composition for an identical object sponsors the inference that the object moved across panels. We capture this in the compositional progression construal CPR1. The graphological structure describes that we can identify two instances of an object (region-a and region-b) as being part of a larger region (e.g. their respective panels). In the conceptual structure, we use ‘G’ this time for the object, to indicate a ‘Goer’: a Material capable of movement/being moved. For the first part of the formula, we have a particular category membership for the Goers in both panels (bound conceptually with  $\alpha$  and  $\beta$ ). Often, the Goer can be inferred to be the same across panels (as in **Figure 9c**), but this construal may also apply to distinct entities (e.g., imagine depictions of different people walking, with each person showing a progressive position along each panel). When there is change between region-a’s place in the first panel and region-b’s place in the second panel, there is the probability that we infer this object (G) went from one place (its position in panel 1) to a different one (its position in panel 2). That movement from one place to another is then interpreted as motion across panels.

Applied to **Figure 9c**, the ball is token-identical across panels ( $\alpha$  and  $\beta$ ) and forms a smaller region within panel 1 and 2 (region-a and region-b respectively). Since the coordinates of the ball in the first panel imply a higher positioning relative to the figure than the coordinates of the ball in the second panel, there is a difference between the positioning of region-a and region-b, and we infer that the ball moved. We omit them here for simplicity, but this coordinate information could be elaborated within the panels’ graphological structure. The inference here tells us that the object went from its first location (the position of graphic region-a) to a different location

(the position of region-b). Hence, we can deduce an object's path of motion based on its relative positioning across panels.

Compositional Progression construal 1 (CPR1):

GS:	$/\rho \dots / \rho X^a / \dots / 1$	$/\rho \dots / \rho X^b / \dots / 2$
CS:	$[G]^a_1$	$[G]^b_2$
CPR1:	$p([\text{Mat./ Sit./ Space } \text{CATEGORY-MEMBERSHIP}(\alpha, \beta)] \mid [\text{Situation } \text{GO}(G^{\alpha,\beta}, [\text{Path}$ $\text{FROM}([\text{Place } F(G^a)]), \text{TO}([\text{Place } F(G^b)])])])$ $p([\text{Mat./ Sit./ Space } a \Delta b > (+\epsilon) 0] \mid [\text{Situation } \text{GO}(G^{\alpha,\beta}, [\text{Path } \text{FROM}([\text{Place } F(G^a)]), \text{TO}([\text{Place}$ $F(G^b)])])])$	

Yet another aspect that may support inferences of movement across panels is the addition and/or omission of objects. In **Figure 7**, we saw that new characters appeared in panel 2, which may elicit the inference that these characters came running from just outside the panel border towards the events visible within the scene. Likewise, when the character on the right does not reappear from panel 2 to 3, readers may assume that this character fell down the platform and moved out of sight. These two inferences are captured in our quantity progression construals: QPR1 (for addition) and QPR2 (for omission). To explain how an addition of a character may imply movement, QPR1 describes a presumed relation between two elements of Space, indicated in the formula with L and L' (with L referring to Location). When a difference between regions corresponds to Materials, it may appear that Material A was introduced as a new entity. Since the narrative remains in a single location and a new input emerges, a subsequent inference follows: Entity A moved from an unknown location (indicated with ?) to a location visible within region-b (indicated with X). In **Figure 7**, the similarity across Spaces in all three panels suggests this is the same location throughout. The added character is then assumed to move from an off-panel location in panel 1 to a spot within the panel border in panel 2.

The reverse interpretation applies to QPR2. When the Spatial locations L and L' are related and an object is replaced by a lack of object, we may construe the object to have disappeared. This interpretation sponsors the subsequent inference that the object A goes from a visible place (region-a) to an unknown, off-panel place (?). From panel 2 to 3 in **Figure 7**, the character disappears, eliciting the idea that they moved from their depicted place in panel 2 to an unknown location in panel 3.

Quantity Progression construal 1-2 (QPR1-2):

GS1:	$/_{\rho} 0 /^a_1$	$/_{\rho} X \dots n /^b_2$
CS1:	$[0]^a_1$	$[A]^b_2$
QPR1 (addition):	$p([\text{Space TOKEN/TYPE-IDENTICAL}(L, L')] \mid [\text{State BE}(\alpha) \Delta [\text{State BE}(\beta)]^y])$ $p([\text{Material } a \Delta b \neq 0] \mid [\text{State BE}(\alpha) \Delta [\text{State BE}(\beta)]^y])$ $p([\gamma] \mid [\text{Situation GO}(A, [\text{Path FROM}([\text{Place F}(\?))], \text{TO}([\text{Place F}(X)]))]))$	
GS2:	$/_{\rho} X \dots n /^a_1$	$/_{\rho} 0 /^b_2$
CS2:	$[A]^a_1$	$[0]^b_2$
QPR2 (omission):	$p([\text{Space TOKEN/TYPE-IDENTICAL}(L, L')] \mid [\text{State BE}(\alpha) \Delta [\text{State BE}(\beta)]^y])$ $p([\text{Material } a \Delta b \neq 0] \mid [\text{State BE}(\alpha) \Delta [\text{State BE}(\beta)]^y])$ $p([\gamma] \mid [\text{Situation GO}(A, [\text{Path FROM}([\text{Place F}(X)]), \text{TO}([\text{Place F}(\?))]))])$	

## 5.2 Static construals (ST)

Conceivably, when there are inferences of progression, comparisons may also elicit inferences of a lack of movement. Consider **Figure 4a**, where the panel pair showed two identical figures against a white background. This continuity allowed the inference that the two figures were the same character shown twice. The lack of any differences across panels may also imply we are shown the same moment twice. Similarly, **Figure 5**, which showed part of a person's braid in the first panel and the entirety of the person in the next, may elicit the same interpretation. As the panel is just zooming out on the scene, a reader may feel like they are viewing a single state. In contrast, **Figure 9b** showed different figures in different locations, involved in seemingly the same action. We described that this discontinuity in Materials and Space paired with complete co-reference for Situation resulted in no apparent temporal/spatial relation between panels. As such, readers may believe these to be separate events, that may even occur simultaneously. Thus, while complete co-reference between Materials and/or Spaces may result in the impression of a singular moment, lack of co-reference can lead to the inference that panels depict unrelated events.

We account for these distinct subsequent inferences in the static construals ST1 and ST2. ST1 describes that complete co-referentiality of Materials/Spaces/Situations probably results in the

inference that both instances show the same state. For ST2, the difference between the Materials/Spaces/Situations evokes separate Situations rather than consecutive moments of one action.

Static construal 1-2 (ST1-2):

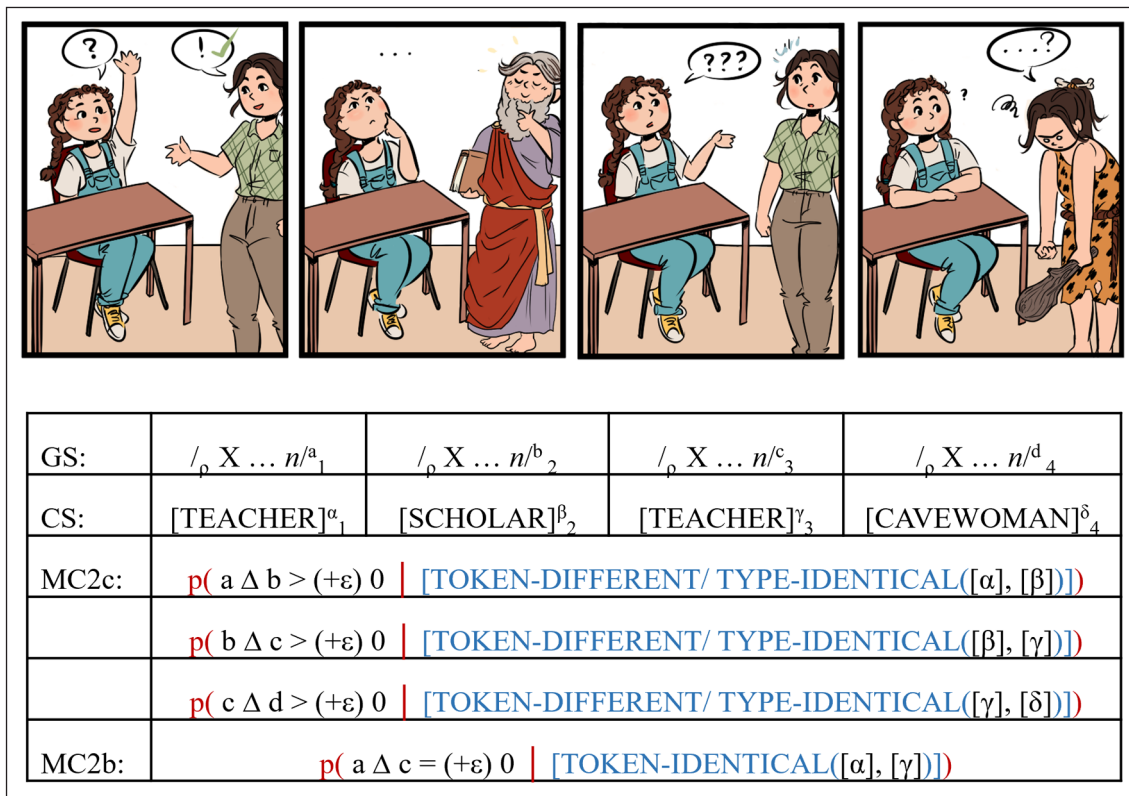
GS:	$/\rho X \dots n/a_1$	$/\rho X \dots n/b_2$
CS:	$[A]^\alpha_1$	$[A]^\beta_2$
ST1:	$p([\text{Mat./ Sit./ Space TOKEN-IDENTICAL}(\alpha, \beta)] \mid [\text{State BE}(\alpha), [\text{State BE}(\beta)]])$	
ST2:	$p([\text{Mat./ Sit./ Space TOKEN-DIFFERENT/ TYPE-IDENTICAL}(\alpha, \beta)] \mid [\text{Situation F}(\alpha), [\text{Situation F}(\beta)]])$	

### 5.3 Construals for meaningful discontinuity

A final set of subsequent inferences related to co-reference comes from discontinuity in visual sequences that can still be meaningfully resolved. Consider **Figure 10**, where in the first panel, a student asks a question to her teacher, who responds with an answer (with the checkmark suggesting her response satisfies the enquiry). While the student thinks on this answer in the second panel, the teacher is replaced by an ancient scholar, stroking their beard happily. In the third panel, the student asks more question(s), and the panel again shows the teacher, looking surprised. In the final panel, the student waits for an answer, seemingly unaware (suggested by the question mark near them and lack of surprise) that the teacher is now replaced by a frustrated cavewoman, who does not seem to know an answer. The evident graphic differences between the teacher and the figures replacing her likely yield the inference that these are distinct tokens (following MC2c), while comparisons made between the teacher in panel 1 and 3 would result in inferences of a same token (following MC2b).

At face value, the teacher in panels 1 and 3 would be physically replaced by different characters in panels 2 and 4, but such incongruity tends to prompt readers to seek an interpretation that explains the discontinuity (Schilperoord 2018). Consequently, readers may explain the scholar and cavewoman as representations of how the woman feels or sees herself (smart as a scholar when she knows the answer vs. uneducated as a cavewoman when she does not). This would mean that those appearance changes do not *actually* occur in the storyworld at that moment. Instead, the discontinuity can be resolved by the interpretation that one set of events represents a character's state, feelings, or thoughts (Abusch & Rooth 2022; Bimpikou 2018; Klomberg et al. 2023; Maier & Bimpikou 2019), supported also by the student appearing unaware of the incongruity (for similar examples, see e.g., Abusch & Rooth (2022) and Maier & Bimpikou (2019)). This interpretation then reconciles the distinct tokens in panel 2 and 4 as referring

to a common identity, namely that of the teacher, who is imagining *herself* with a different appearance.



**Figure 10:** Example showing intentional, meaningful discontinuity. Image © Bien Klomberg.

While this example may appear specific, corpus work investigating 40 comics across four countries showed that 67.5 % included such forms of intentional discontinuity (van der Gouw et al. 2022). Incongruous appearances for the same character may occur when the narrative switches to a character’s (mental) state or thoughts, such as showing someone’s memories, imagination, or dreams, where they may appear significantly younger or in a different shape (Abusch & Rooth 2022; Bimpikou 2018; Klomberg et al. 2023). The events depicted as part of those thoughts are considered “auxiliary events”, that are understood as not actually taking place in that moment or place in the narrative, but instead as visualizations of private thoughts or (mental) experiences that can depict the characters in the story distinctly different from their usual, expected appearance (Klomberg et al. 2023). In **Figure 10**, as we reach the final panel, we can see a pattern unfolding where the teacher in her green blouse would be the expected appearance (reappearing in panel 3), and the scholar and cavewoman are (mental/emotional) reactions to how well the teacher can respond to the student’s questions. These cases support

that co-referentiality does not rely on perceptual or visual cues alone but requires subsequent inferences to account for more complex cases.

We will class these types of inferences as Discontinuity Construals (DC), for which DC1.1 involves a token- or type-different category membership between the relevant regions, and, optionally, the level of change across regions. One or both of these analyses will lead to an inference of what may be going on with the involved characters. There may be an overt ‘Experiencer’ (E) that is the one remembering/imagining/dreaming the auxiliary events (e.g., the teacher in **Figure 10**), but it may also be left implicit who is responsible for the discontinuity in depictions. As described in DC1.1, readers infer a certain action (indicated by the general function F) that the character (the Experiencer, E) would be involved in, which bears consequences for the relation (R) between the two discontinuous units (A and A').

Discontinuity Construal 1.1-1.2 (DC1.1-1.2):

GS:	$/_{\rho} X \dots n^a_1$	$/_{\rho} X \dots n^b_2$
CS:	$[A]^a_1$	$[A']^b_2$
DC1.1	$p([\text{Mat./ Sit./ Space} \text{ TOKEN/ TYPE-DIFFERENT}(\alpha, \beta)] \mid [\text{Situation F(E, [R}(\alpha, \beta)]^{\gamma})])$ $p([\text{Mat./ Sit./ Space} \text{ a } \Delta \text{ b} > (+\epsilon) 0] \mid [\text{Situation F(E, [R}(\alpha, \beta)]^{\gamma})])$	
DC1.2	$\text{If}(\gamma = \text{BE}(\alpha, \beta) \rightarrow [\text{TOKEN-IDENTICAL}([\alpha], [\beta])])$	

Let us apply DC1.1 to **Figure 10**'s first change, where the teacher imagines herself as an ancient scholar to reflect how smart she thinks/feels she is. As the region of the scholar differs significantly from the region of teacher, these adhere to a token-different relation. Comprehenders will aim to resolve this discontinuity and as such, likely deem the scholar to be ‘unreal’ (an auxiliary event), originating from the teacher’s mind. The inference described abstractly in DC1.1 would then be filled in as:

$[\text{Situation IMAGINE}(\text{TEACHER, [BE}(\text{TEACHER, SCHOLAR})]^{\gamma})]$

Namely, the teacher is the Experiencer, filling in the place of E. The function that explains the discontinuity is that the teacher *imagines* (F) a change from her own self to the scholar. This sense of identity (of ‘being’) is the relation (R) between the two disparate appearances (A and A'). The entirety of this inference is indicated with a binding operator  $\gamma$ .

DC1.2 then describes that if this inference ( $\gamma$ ) implies a similar identity between the two units (with  $\alpha$  and  $\beta$  being the same), then we can interpret the two distinct Materials to actually be token-identical. For **Figure 10**, the inference describes that the teacher imagines herself to be a



different person. This links the two appearances as originating from the same source. Moreover, ‘imagination’ denotes a mental visualization, which explains the featural discrepancies in a logical manner. As such, a similar identity can be ascribed to the two distinct regions, leading to the subsequent inference reconciling the two regions as the same conceptual entity (token-identical). The same process would apply to **Figure 10**’s second change (from teacher to cavewoman). The last recognizable change, from the scholar in panel 2 back to the teacher, would then constitute an interruption of the imagination process, when the teacher stops thinking of herself in this way.

We can apply these same construals to cases where it is not entirely clear who initiates the visual discontinuity. A comprehender may disagree with our analysis that the student remains unaware and may see it as ambiguous who imagines these changes in appearance. Moreover, it could also be unclear whether the act that evokes auxiliary events is imagining, remembering, dreaming, etc. Imagine if **Figure 10** included the student falling half-asleep; perhaps then it is ambiguous whether the appearance changes are due to imagining or dreaming. Since the DC1.1 construal has slots that can be filled in, analyses can represent such ambiguity by leaving those slots open:

[<sub>Situation</sub> ?(?, [IS-LIKE(TEACHER, SCHOLAR)]<sup>v</sup>)]

While we can still infer a relation between the two figures, with a comparison being made between the teacher and this scholar, left unspecified (?) are the function F that ties together this relation between the teacher and scholar (i.e., imagination, dreaming, etc.) and the Experiencer of this relation. Despite these unspecified entries, an analogy remains, which may still lead us to DC1.2, where based on an inferred similarity between the discontinuous forms, we can interpret that these forms represent a single entity. This supports how visual discontinuity can be resolved meaningfully, even in cases without overt Experiencers.

Some readers may interpret the change from the teacher to the scholar (and later, from teacher to cavewoman) as a straightforward switch in the narrative between two figures, perhaps simply a comparison between them that does not unite the figures as similar in identity to one another. That interpretation evokes then only DC1.1, which describes a relation between entities, and not DC1.2, which reconciles the entities as referring to the same entity. Similarly, one may interpret some of these cases of visual discontinuity as actual transformations between entities, where one entity physically changes into the other (e.g., through magic or shapeshifting powers). If the entity’s identity is retained in their new form, one could still apply DC1.2, but when this is not the case, this would warrant only DC1.1, since we now have two separate entities. The function F then becomes the physical act of ‘TRANSFORM’. Below we outline this inference as applied to **Figure 10**.

[<sub>Situation</sub> TRANSFORM(?, [GO(TEACHER, [<sub>Path</sub> FROM(TEACHER), TO(SCHOLAR)])])]

This Discontinuity construal thus includes two parts, where it depends on the interpretation of the reader (and potentially the context of the narrative, e.g. if magic/shapeshifting is allowed) whether both are evoked. This way, we can account for the same example eliciting different interpretations from different readers.

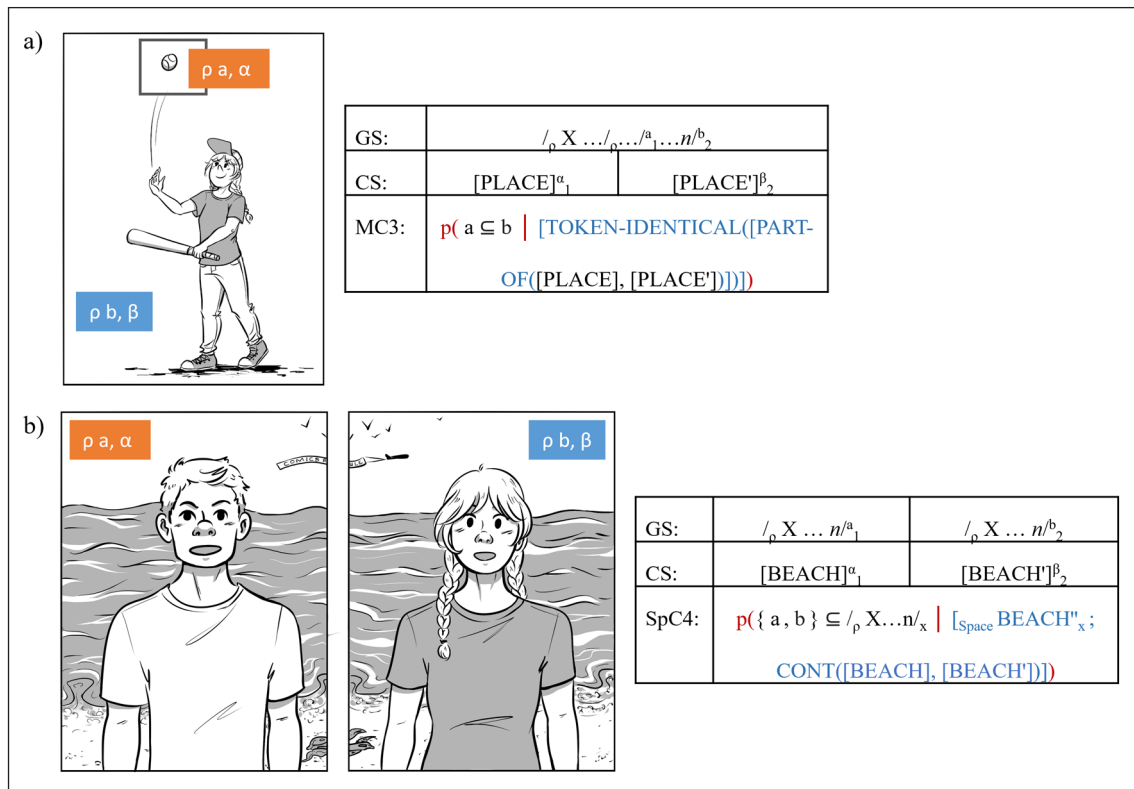
## 6 More complex paneling structures

So far, we have demonstrated the use of our model for analyzing multiple aspects of an image and for accounting for the variety of inferences that may arise from the juxtaposition of images. Next, we show how our model applies not only to juxtapositions of panels, but also to more complex framing structures. Here we outline the co-referentiality inferences related to two more complex paneling structures: inset panels (**Figure 11a**) and divisional panels (**Figure 11b**) (Cohn 2014). Inset panels (see region-a in **Figure 11a**) are panels encompassed by a larger, “dominant” panel (region-b in **Figure 11a**), which affects the graphological structure. Instead of two separate panel, the GS would thus be:

GS:	$l_p X \dots l_p \dots l_1 \dots n^b_2$
-----	---

When insets show a portion of their dominant panel, our Material zoom-in constraint (MC3) can be effectively applied to Space. This constraint then shows that when the contents of one region (the dominant panel in **11a**) encompasses that of another region (the inset panel in **11a**), there is the probability that we interpret the second region’s spatial representation ( $\alpha$  in **11a**) to be a part of the first region’s overall spatial representation ( $\beta$  in **11a**). In other words, we presume the inset panel’s spatial location is identical to the dominant panel’s Space.

Divisional panels are addressed in SpC4. These are panels which use image constancy where a single depiction is divided into different panels, or alternatively, where content from each panel connects to create the impression of a larger image (see **Figure 11b**). In **Figure 11b** both panels build up to a larger scene due to the continuity of elements across panel borders (e.g., the plane’s banner, the seaweed, the waves). SpC4 therefore proposes that when regions corresponding to panels (region-a and region-b) together constitute ( $\subseteq$ ) a larger region ‘L’, there is the probability that L is inferred as a larger Space containing (the function “CONT” for contain) both Spaces A and A’ (corresponding to region-a and region-b). In simpler terms, when the graphics across panels form a continuous environment, we may interpret the panels as components of a larger, holistic Space. In **Figure 11b**, the two beach environments together constitute one larger beach scene.



**Figure 11:** Examples showing the more complex framing structures of a) inset panels and b) divisional panels.

Spatial Constraint 3 (SpC3):

GS:	$/_{\rho} X \dots n^{\alpha}_1$	$/_{\rho} X \dots n^{\beta}_2$
CS:	$[A]^{\alpha}_1$	$[A]^{\beta}_2$
SpC3:	$p(\{a, b\} \subseteq /_{\rho} X \dots n^x \mid [_{Space} L_x; CONT([\alpha], [\beta])])$	

While we specify these construals for Space, paneling can also be applied to Materials. For example, a divisional panel could partition panels across the parts of a human body, or an inset could focus on one body part within a dominant panel of human figure. In these cases, the formalisms would remain the same, only the ontological categories would change from Space to Material.

## 7 Discussion

Overall, we presented a model of constraints that describes varying relations of co-reference based on graphics and their corresponding conceptual structure. We aimed to describe what

comprehenders would need to learn to understand continuity and co-reference across images, in response to empirical and developmental studies suggesting the need for some level of proficiency. The resulting model accounts for incremental and/or partial change across three superordinate categories of meaning, as well as subsequent inferences that may arise based on the (dis)continuity across those categories. Those inferences include temporal and spatial relations across panels and the reconciliation of a specific structure of visual discontinuity. Such spatiotemporal relations connect to analyses from prior theories (Bateman & Wildfeuer 2014; Maier & Bimpikou 2019; Saraceni 2016; Schlöder & Altshuler 2023; Stainbrook 2016), but while those studies seemed to question co-reference by examining how events across panels relate to each other, the current paper instead aimed to examine how *graphic lines* motivated comprehenders to infer (no) co-reference. The interactions between co-reference analyses across different dimensions were then found to provide a base to explain how subsequent (spatiotemporal) inferences are actually achieved. We speculate these spatiotemporal relations to then interface with more complex panel relations, such as causality or other discourse relations.

All constraints of the three categories of meaning are set up in a similar way (with both top-down and bottom-up processes), but the effect of these constraints may differ based on expectancies of how often information in that category tends to change. Corpus research has shown that the situational aspects of time, spatial location, and characters have different patterns of continuity in visual sequences, with generally few changes across time and spatial locations and relatively many across characters (Klomberg et al. 2022). This suggests a gradation in top-down likelihoods (laid out in our constraint 1 types), with stronger predictions for Spatial and Situational co-reference across panels than Material co-reference. Furthermore, patterns of situational (dis)continuity differ cross-culturally, with American and European comics maintaining greater temporal and character continuity and less spatial continuity than Asian comics (Klomberg et al. 2022). Therefore, the probabilities articulated by these constraints may hold different weights depending on culturally entrenched patterns. Moreover, due to these top-down likelihoods, token- or type-identical category membership may also have slightly different effects depending on the conceptual category. For Materials and Spaces, complete co-referentiality in the form of token-identical units align with top-down predictions, but for Situations, panels are generally predicted to depict *related* Situations rather than identical ones. Therefore, complete Situational co-referentiality across panels (which would be token-identical) may feel less continuous to readers than distinct parts of a larger movement (type-identical).

We acknowledge here that actual reading of visual narratives likely involves more predictions than we have yet described, based for example on characters' prior actions, motives, or even conventions of the genre, etc. (Cohn 2020) This again aligns with the meanings in spoken or signed languages, where circumstances of utterances may grant interpretations not present in the linguistic material alone. While formalizing such world knowledge poses a challenge no

matter the modality, our formalized framework is capable of accounting for visual meaning-making in both typical examples and outliers, and affords modification to account for those various circumstances. For instance, our Discontinuity construal accounts for narratives with and without overt Experiencers and interpretations of mental states vs. physical transformations, which may become possible when the story allows magic/shapeshifting. In addition, if genre is a predictor of certain co-referential constructions (e.g., horror genres more often show meaningful discontinuity), this may simply mean that readers of that genre alter the range of likelihood as expressed by our probability statements and/or have specific constructions encoded within the registers associated with those genres (e.g. readers of those genres have a lower threshold for applying DC1.2 than readers of other genres). This would be similar to our predictions of comprehension for cross-linguistic differences, where familiarity of comics with particular patterns have been shown to modulate their comprehension (Cohn 2020).

Indeed, these proposed constraints on co-reference connect to broader processes of cognition. Our forward predictions outlined in the constraint 1 types correspond to expectancies reflected in the forward-looking processes involved in processing visual sequences (Cohn 2020), which would have to be acquired through exposure and practice with visual narratives. The constraints outlined in the 2 to 3 types would reflect “mapping” updating processes (Gernsbacher 1985; 1997; Loschky et al. 2020). These constraints may potentially reflect the amount of updating that is necessary, with increasing discontinuity for constraint types 2 implying increasing cognitive efforts. Subsequent inferences and construals would pose greater challenges to cognition, and as such more align with “shifting”, a greater updating process warranting inference or creation of new mental models (Gernsbacher 1985; 1997; Loschky et al. 2020). Theories of panel transitions (Gavaler & Beavers 2018; McCloud 1993; Saraceni 2016) or discourse/coherence relations (Bateman & Wildfeuer 2014; Maier & Bimpikou 2019; Saraceni 2016; Schlöder & Altshuler 2023; Stainbrook 2016) do not capture such nuances between types of co-referentiality processes, often lack any distinction between inferencing and mapping, and do not seem to support forward predictions. However, based on results from psychological studies, these appear essential mechanisms to consider for visual comprehension.

The varying updating efforts of these mechanisms raise the question of how taxing these processes would be throughout cognition. While establishing co-referentiality may appear effortless for many readers, our model suggests the complexity that lies beneath the surface, even for just two juxtaposed panels. For example, varying continuity assessments can persist across multiple dimensions, the combination of which may again prompt varying subsequent inferences. Some work has claimed readers establish meaningful relationships potentially between every panel in a book (Gavaler & Beavers 2018; Groensteen 2007). Even if semantic relationships between panels were binary and exclusive, such “promiscuous transitions” would be incredibly demanding for processing (Cohn 2010), let alone with multiple co-occurring

co-referential connections, as argued in these constraints. That is, coherence relationships like these are insufficient on their own to account for sequential image comprehension, and in order to negotiate potential distance dependencies between panels (Cohn 2013; 2014), additional structure(s) are required. To this end, a parallel narrative grammar has been posited to help guide this semantic processing (Cohn 2013; 2014; Cohn et al. 2014), just as the syntax of speech interfaces with its semantics (Jackendoff 1983; 2002). Such a system would provide segmentation through recursive embedding, along with categorical roles differentiating the saliency of panels, all of which would alleviate demands on working memory for establishing co-reference.

Despite these structures involved, the complexity of visual co-referentiality is not often readily recognized throughout the comprehension process (Cohn 2020), perhaps leading to the beliefs that the understanding of visual sequences is universal and/or can be ceded to pragmatics or visual perception (Abusch 2012; Bateman & Wildfeuer 2014; Gavaler & Beavers 2018; Maier & Bimpikou 2019). One reason that this process may appear deceptively easy to readers (and researchers) may be due to their own fluency in visual language, which has been shown to affect comprehension of visual sequences (Cohn 2020; Lichtenberg et al. 2022). Such fluency may support why some perceive co-referentiality as a matter of pragmatics, as the learned constraints then appear to become obvious, seemingly inherent knowledge. With not all readers and ages able to construe continuity in sequential images (Cohn 2020), however, such constraints must be acquired through exposure to visual languages. Without such exposure, principles underlying continuity remain inaccessible, which accounts for why some readers (presumably still with access to perception, general world knowledge, and pragmatic principles) may fail to construe visual co-reference. For instance, encoding of the MC2b constraint (i.e., co-reference across figures despite minimal changes) would account for how fluent readers can interpret similarly drawn figures with progressively longer beards as *one* person whose beard grows. Unfamiliarity with this constraint then explains why non-fluent readers may see these as “brothers”; they perceive the similarity across figures but fail to connect those similar depictions as co-referential.

In line with more recent approaches in linguistics (Cohn 2020; Patel-Grosz et al., In press.), we argue that studying meaning-making across modalities yields insights both for individual modalities, and for studies of our language faculty as a whole. Our approach emphasizes that different modalities achieve similar conceptual inferences, despite the distinct mechanisms that underlie form-meaning interfaces. The complexity of those interfaces further substantiates visual narratives as rich, non-transparent communicative systems, just like spoken and signed languages. While different modalities are often perceived as distinct and highly varying, they may maintain overlapping structures and resources. This may result in similarities between modalities persisting on a more abstract, structural level, manifesting according to the affordances of each expression. For instance, verbal language has particular conventionalized expressions as foundations for co-reference, while visual language then may rely on (probability) constraints, similarly learned

by comprehenders. Better understanding of how such differences operate within and across modalities can thus further contribute to how they combine in multimodal interactions, and how they relate in cognition.

In conclusion, our model of co-reference in visual sequences not only describes that co-reference does occur but seeks to describe how we create the mapping between a graphic representation and meaning in the first place. This results in formalizations of co-reference in visual language, providing insights into the affordances of graphic communication, and how visual expressions, while differing from other language systems in form, can accomplish the same meaning. These formalizations would need to be acquired by comprehenders through exposure with visual language, which explains how co-reference can be construed but also why it might not be in absence of familiarity with these principles. In line with psychological literature on visual sequence comprehension and fluency, our model thus demonstrates the complexity of continuity, co-reference, and inference in visual sequencing.

---

## Funding information

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 850975).

## Competing interests

The authors have no competing interests to declare.

---

## References

- Abusch, Dorit. 2012. Applying discourse semantics and pragmatics to co-reference in picture sequences. In *Proceedings of Sinn und Bedeutung*, Vol. 17.
- Abusch, Dorit & Rooth, Mats. 2022. Temporal and intensional pictorial conflation. In *Proceedings of Sinn und Bedeutung*, Vol. 26.
- Bateman, John & Wildfeuer, Janina. 2014. A multimodal discourse theory of visual narrative. *Journal of Pragmatics* 74. 180–208. DOI: <https://doi.org/10.1016/j.pragma.2014.10.001>
- Biederman, Irving. 1987. Recognition-by-components: A theory of human image understanding. *Psychological Review* 94(2). 115–147. DOI: <https://doi.org/10.1037/0033-295X.94.2.115>
- Bimpikou, Sofia. 2018. Perspective blending in graphic media. In *Proceedings of the ESSLLI 2018 Student Session*, 245–257. Sofia, Bulgaria: ESSLLI. Retrieved from <https://core.ac.uk/download/pdf/232523553.pdf#page=246>
- Braithwaite, Jean & Mikkonen, Kai. 2022. Figural solidarity: Grappling with meaning in comics. *Comicalités. Études de Culture Graphique*. DOI: <https://doi.org/10.4000/comicalites.7729>
- Chomsky, Noam. 1980. On binding. *Linguistic Inquiry* 11(1). 1–46.
- Cohn, Neil. 2010. The limits of time and transitions: Challenges to theories of sequential image comprehension. *Studies in Comics* 1(1). 127–147. DOI: <https://doi.org/10.1386/stic.1.1.127/1>
- Cohn, Neil. 2012. Explaining ‘I can’t draw’: Parallels between the structure and development of language and drawing. *Human Development* 55(4). 167–192. DOI: <https://doi.org/10.1159/000341842>
- Cohn, Neil. 2013. *The visual language of comics: Introduction to the structure and cognition of sequential images*. Bloomsbury Academic.
- Cohn, Neil. 2014. The architecture of visual narrative comprehension: The interaction of narrative structure and page layout in understanding comics. *Frontiers in Psychology* 5. DOI: <https://doi.org/10.3389/fpsyg.2014.00680>
- Cohn, Neil. 2020. *Who understands comics? Questioning the universality of visual language comprehension*. London: Bloomsbury Academic. DOI: <https://doi.org/10.5040/9781350156074>
- Cohn, Neil & Jackendoff, Ray & Holcomb, Phillip & Kuperberg, Gina. 2014. The grammar of visual narrative: Neural evidence for constituent structure in sequential image comprehension. *Neuropsychologia* 64. 63–70. DOI: <https://doi.org/10.1016/j.neuropsychologia.2014.09.018>



- Cohn, Neil & Schilperoord, Joost. 2022. Reimagining language. *Cognitive Science* 46(7). e13164. DOI: <https://doi.org/10.1111/cogs.13174>
- Coopmans, Cas W. & Cohn, Neil. 2022. An electrophysiological investigation of co-referential processes in visual narrative comprehension. *Neuropsychologia* 172. DOI: <https://doi.org/10.1016/j.neuropsychologia.2022.108253>
- Frederiksen, Anne Therese & Mayberry, Rachel I. 2022. Pronoun production and comprehension in American Sign Language: The interaction of space, grammar, and semantics. *Language, Cognition and Neuroscience* 37(1). 80–102. DOI: <https://doi.org/10.1080/23273798.2021.1968013>
- Gavaler, Chris & Beavers, Leigh Ann. 2018. Clarifying closure. *Journal of Graphic Novels and Comics* 11(2). 182–210. DOI: <https://doi.org/10.1080/21504857.2018.1540441>
- Gernsbacher, Morton Ann. 1985. Surface information loss in comprehension. *Cognitive Psychology* 17(3). 324–363. DOI: [https://doi.org/10.1016/0010-0285\(85\)90012-X](https://doi.org/10.1016/0010-0285(85)90012-X)
- Gernsbacher, Morton Ann. 1997. Coherence cues mapping during comprehension. In J. Costermans & M. Fayol (eds.), *Processing interclausal relationships in the production and comprehension of text*, 3–21. Hillsdale, NJ: Erlbaum.
- Gordon, Peter C. & Hendrick, Randall. 1997. Intuitive knowledge of linguistic co-reference. *Cognition* 62(3). 325–370. DOI: [https://doi.org/10.1016/S0010-0277\(96\)00788-3](https://doi.org/10.1016/S0010-0277(96)00788-3)
- Graesser, Arthur C. & Millis, Keith K. & Zwaan, Rolf A. 1997. Discourse comprehension. *Annual Review of Psychology* 48. 163–89. DOI: <https://doi.org/10.1146/annurev.psych.48.1.163>
- Groensteen, Thierry. 2007. *The system of comics*. (B. Beaty & N. Nguyen, Trans.) 1st ed. University press of Mississippi.
- Gruber, Jeffrey S. 1965. *Studies in lexical relations*. Massachusetts Institute of Technology dissertation.
- Jackendoff, Ray. 1983. *Semantics and cognition*. MIT Press.
- Jackendoff, Ray. 2002. *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780198270126.001.0001>
- Jackendoff, Ray. 2010. *Meaning and the lexicon: The parallel architecture 1975-2010*. Oxford University Press.
- Klomberg, Bien & Hacımusaoglu, Irmak & Cohn, Neil. 2022. Running through the who, where, and when: A cross-cultural analysis of situational changes in comics. *Discourse Processes* 59(9). 1–16. DOI: <https://doi.org/10.1080/0163853X.2022.2106402>
- Klomberg, Bien & Schilperoord, Joost & Cohn, Neil. 2023. *Constructing domains in visual narratives: Structural patterns of incongruity resolution*. Manuscript in press, Department of Communication and Cognition, Tilburg University.
- Le-Hoa Vō, Melissa. 2021. The meaning and structure of scenes. *Vision Research* 181. 10–20. DOI: <https://doi.org/10.1016/j.visres.2020.11.003>
- Lerdahl, Fred & Jackendoff, Ray. 1983. *A Generative theory of tonal music*. The MIT Press.
- Lichtenberg, Lenneke Doris & Hacımusaoglu, Irmak & Klomberg, Bien & Schilperoord, Joost & Cohn, Neil. 2022. *A closer look to the monkey emoji debate: Assessing the continuity constraint*

using emoji sequences. Manuscript in preparation, Department of Communication and Cognition, Tilburg University.

Loschky, Lester C. & Larson, Adam M. & Smith, Tim J. & Magliano, Joseph P. 2020. The scene perception & event comprehension theory (SPECT) applied to visual narratives. *Topics in Cognitive Science* 12(1). 311–351. DOI: <https://doi.org/10.1111/tops.12455>

Maier, Emar & Bimpikou, Sofia. 2019. Shifting perspectives in pictorial narratives. In *Proceedings of Sinn und Bedeutung*.

Mandler, J.M. 2004. *The foundations of mind: Origins of conceptual thought*. Oxford University Press. DOI: <https://doi.org/10.1111/j.1467-7687.2004.00369.x>

Marr, David. 2010. *Vision: A computational investigation into the human representation and processing of visual information*. London: The MIT Press. DOI: <https://doi.org/10.7551/mitpress/9780262514620.001.0001>

McCloud, Scott. 1993. *Understanding comics: The invisible art*. Northampton: Kitchen Sink Press.

McCloud, Scott. 2000. *Reinventing comics: How imagination and technology are revolutionizing an art form*. Harper Collins Publishers.

Napoli, Donna Jo & Leeson, Lorraine. 2020. Visuo-spatial construals that aid in understanding activity in visual-centred narrative. *Language, Cognition and Neuroscience* 35(4). 440–465. DOI: <https://doi.org/10.1080/23273798.2020.1744672>

Patel-Grosz, Pritty & Mascarenhas, Salvador & Chemla, Emmanuel & Schlenker, Philippe. In press. Super linguistics: An introduction \*. *Linguistics & Philosophy*.

Sanders, Ted J.M. & Gernsbacher, Morton Ann. 2004. Accessibility in text and discourse processing. *Discourse Processes* 37(2). 79–89. DOI: [https://doi.org/10.1207/s15326950dp3702\\_1](https://doi.org/10.1207/s15326950dp3702_1)

Saraceni, Mario. 2016. Relatedness: Aspects of textual connectivity in comics. In Neil Cohn (ed.), *The Visual Narrative Reader*, 115–128. Bloomsbury Publishing. DOI: <https://doi.org/10.5040/9781474283670.ch-005>

Schilperoord, Joost. 2018. Ways with pictures: Visual incongruities and metaphor. In Gerard Steen (ed.), *Visual Metaphor: Structure and process*, 11–47. Amsterdam: John Benjamins Publishing Company. DOI: <https://doi.org/10.1075/celcr.18.02sch>

Schlöder, Julian J. & Altshuler, Daniel. 2023. Super pragmatics of (linguistic-)pictorial discourse. *Linguistics and Philosophy*. DOI: <https://doi.org/10.1007/s10988-022-09374-x>

Stainbrook, Eric. 2016. A little cohesion between friends; or, we're just exploring our textuality: Reconciling cohesion in written language and visual language. In Neil Cohn (ed.), *The Visual Narrative Reader*, 129–156. Bloomsbury Publishing. DOI: <https://doi.org/10.5040/9781474283670.ch-006>

Talmy, Leonard. 2003. *Towards a cognitive semantics: Typology and process in concept structuring*. The MIT Press.

Tseng, Chiao-I. & Bateman, John. 2018. Cohesion in comics and graphic novels: An empirical comparative approach to transmedia adaptation in *City of Glass*. *Adaptation* 11(2). 122–143. DOI: <https://doi.org/10.1093/adaptation/apx027>

van der Gouw, Sharitha & Klomberg, Bien & Cohn, Neil. 2022. *Visual storytelling: A corpus analysis of domain usage in American and Japanese comic styles*. Tilburg, the Netherlands: Tilburg University dissertation.

Willats, John. 1997. *Art and representation: New principles in the analysis of pictures*. Princeton University Press.

Willats, John. 2005. *Making sense of children's drawings*. New Jersey: Lawrence Erlbaum Associates. DOI: <https://doi.org/10.4324/9781410613561>

