Narrative Patterns and Memory:

Do we retain visual narrative structures of sequences as we do with the meaning?

Renel van de Wal

ANR: 135730

Thesis supervisors:

Dr. N.T. Cohn

Irmak Hacımusaoğlu M.Sc.

Second reader:

Dr. M. Faber

Master thesis

Master Communication and Information Sciences

Track: Communication and Cognition

School of Humanities and Digital Sciences

June 2021

Tilburg University, Tilburg

Abstract

When processing verbal discourse, we forget the surface information quickly and store the gist in our memory. According to discourse theories the information that we store, gets stored in the form of mental models (Van Dijk & Kintsch, 1983). Whether this is the case for visual narratives, which is another form of communication, has not been studied so far. Previous research has been done on how long we remember images, but those studies focused on physical dimensions of the images and not entire sequences. Besides, according to the theory of Visual Narrative Grammar, the basic form of image sequencings is the canonical narrative schema, which can be altered with different patterns (Cohn, 2018). This study seeks evidence for whether we forget these surface narrative structures the same way we do with verbal and written discourse and also aims to find out what influence the different narrative patterns might have on the retainment of surface structures. The experiment of this study was done in 2 sessions. First, participants were given sequences with different narrative structures and had to rate the comprehensibility of the sequences. One week later, participants received sequences that were either the same, had a different structure, or were completely different yet semantically related sequences. Participants had to answer if they had seen the sequences before the first session and how confident they were about their answer. The results of their recognition showed that the narrative structure of visual narrative sequences does not get remembered and we only remember the gist. This means that there is no effect of pattern on retainment since there is no memory for the narrative structure. These results give reason to believe that the theory about the Situation Model is domain-general and also applies to visual discourse.

Keywords: visual language, visual narrative sequences, memory

Introduction

When receiving new information, whether it is through written or verbal discourse, multiple cognitive processes happen before the information is comprehended and stored in our long-term memory (Zwaan & Radvansky, 1998). When someone receives information via a discourse, first they process the surface level, which is the exact wording of the text. Then it is processed on the text-base level, which contains the meaning of the text and forms a network of propositions. Lastly, there is the mental model level (also called the Situation Model), where we build constructions of the events and actions that happen in a discourse (Van Dijk & Kintsch, 1983). When a reader goes through these cognitive processes, with each step, information loss occurs until the gist is retained in our memory.

Not only do written and verbal discourse present meaningful information, similarly, so do narrative sequential images. However, less is known about the cognitive processes of which visual narratives are comprehended and stored in our long-term memory. Previous research shows that the surface level of images stays in our memory longer than the surface level of written and verbal discourse, but eventually also fades away (Baggett, 1975). We assume cognitive processes allow us to store narrative visual sequential in mental models, but it is not clear how precisely. Thus, this study aims to examine the relationship between visual narrative structure and mental model construction in memory.

In like manner to grammar dictating how to form sentences, a similar system contains rules for organizing visual narrative sequences. Visual Narrative Grammar describes the system of rules that guide visual narrative sequences (Cohn, 2018). The most basic sequencing pattern according to the Visual Narrative Grammar is the canonical narrative schema (Cohn, 2013), but a visual narrative sequence can take on other forms. A Conjunction, Alternation, or Refiner Displacement can be added into the sequence to make it more complex. It is unclear whether and if so, how the different narrative patterns are of influence on our ability to store information presented in them in our mental models.

A few studies have looked at the process of memorizing information from visual narratives, but so far none have looked at the relationship between the narrative structure of whole sequences and the Situation Model. Therefore, this study examines if surface structure of images is retained in our memory, the relationship between the different narrative structures of sequences, and their effect on the creation of mental models.

Mental Models of Verbal Discourse

Mental Models

When receiving new information through written or verbal discourse, this novel information needs to be incorporated into our existing knowledge to understand the discourse. Our existing knowledge is stored in mental models. Mental models are the knowledge someone has about a particular domain (Hemforth & Konieczny, 2006). In these mental models, we update our current information with the new information we receive. Updating our existing mental models with the events and actions presented in the discourse is what enables us to process language (Zwaan & Radvansky, 1998). The updated information in our mental models is not the literal information a reader has received from the discourse. According to discourse theories, a piece of discourse is processed on three cognitive levels before it reaches the form in which it gets stored in our mental model (Van Dijk & Kintsch, 1983).

The first level proposed by Van Dijk and Kintsch (1983) is the surface structure. The surface structure is the exact wording that is used in a discourse. At this stage, no connections are made between the individual words yet. This information about the exact surface form fades from our memory very quickly (Gernsbacher, 1985). The text-base level is the second level of discourse processing. The text-base level contains the meaning of the text, which is turned into a network of connected propositions. Propositions are the smallest meaningful units that can be assigned. This level contains the internal meaning of the text. As a result, the information of the text-base level stays a little longer in our memory than the surface-level information (Gernsbacher, 1985). The third level is the mental model (Van Dijk and Kintsch use the term situation model to describe a certain type of mental model constructed in discourse understanding, but for clarity this text will use the term mental model). In this last stage of text comprehension, the textual information is taken and integrated into the prior knowledge (i.e., information that is already stored in our mental models). The new textual information is used to update our current, existing mental model by making inferences between the new information and our already existing knowledge. The information that is eventually retained in our mental models stays there long-term (Gernsbacher, 1985).

Remembering the Gist in Mental Models

This theory outlines that the information a reader remembers from discourse is not the surface structure. Readers do not remember word-by-word what was mentioned in the discourse. Rather, what gets stored in our mental models is just the gist of the given discourse (Zwaan &

Radvansky, 1998). The study by Sachs (1967) demonstrated this phenomenon. In the experiment, subjects listened to a story that included either sentence A or B:

A. He sent a letter about it to Galileo, the great Italian scientist.

B. A letter about it was sent to Galileo, the great Italian scientist.

Afterward, participants had to say whether they heard sentence A or B. If they were asked this question immediately after hearing the target sentence, their ability to pick the correct sentence was about 90%. If they were asked to pick which sentence they had heard after hearing an additional 80 syllables, the accuracy was about chance level. That is to say, surface structure is forgotten rapidly after reading it.

The results of the experiment by Sachs' (1967) suggest that surface representation is forgotten much faster than the gist, which gets remembered as a mental model. When presented with a discourse, different readers take different pieces of information as the gist to remember and save in their mental model (Kintsch & Van Dijk, 1978). For this reason, two people who read the same text with different goals in mind can end up with different mental models. What is remembered as a gist is based on a few factors, such as the reading goal. If you are reading a text to find specific information, you will only save the information you were looking for, as opposed to someone who starts reading a text without a goal and might remember the general plot as the gist (Kintsch & Van Dijk, 1978).

Event Indexing Model

Zwaan, Langston and Graesser (1995) further describe how readers construct situations described in discourse to store in mental models. According to the Event Indexing Model, when a reader comprehends a story, they construct representations of the different entities in that discourse: the characters, events, states, goals, and actions. With each new event or action that takes place, the reader needs to update their mental model on a number of indices (Zwaan, Langston & Graesser, 1995).

There are five types of indices that are constructed when processing a story: temporality, spatiality, protagonist, causality, and intentionality. Readers index when and where the events in a story happen, who the characters are, the causal status with regards to prior events, and the relatedness to the protagonist's goals. When there is a change in the status of any of these indices, the mental model of the reader gets updated so it reflects the new status of the discourse. The updating of mental models is a constant and ongoing process as the reader furthers in a discourse (Kintsch & Van Dijk, 1978).

Mental Models of Visual Narratives

The Mental Model Theory applied to Visual Narratives

Most research on how mental models are constructed and updated with information has come from studies of verbal discourse, but images are also used when communicating. As with text, information gets taken from images and visual narratives which gets processed to build our mental model with. Previous research has demonstrated that with verbal discourse, the gist persists in our memory and the surface structure does not (Van Dijk & Kintsch, 1983). There is reason to believe visual narratives are comprehended similarly. There are comparisons between verbal and visual discourse in research. According to Cohn and Magliano (2020b), there are similar hypotheses on how models of our comprehension progress for both visual and verbal discourse. When reading sequential images, we gather information from events in individual images and convert these into broader structures which are connected. We are also constantly updating our mental models with new information from images. The same processes happen with verbal discourse (Gernsbacher, 1985). These are just two examples of how similar information extraction and comprehension are for both verbal and visual discourse. Previous research gives us reason to believe our memory operates the same for both types of discourse, but there is no confirmation for it. Even though it is possible there would be an overlap, the memory properties have never been tested.

The study by Cohn and Magliano (2020b) is one of multiple that shows the close similarities in discourse comprehension for verbal and visual discourse. Baggett (1975) did research on our image comprehension. With images, there are two levels at which the image has to be processed for it to be comprehended (Baggett, 1975). The first one is the surface level, at which readers take information from the image they see. Surface information is what readers can take from an image directly as it is shown to them. The research demonstrated that the memory for the surface level of images is stronger than for the surface level for written and verbal discourse. The outcome of Baggett's study (1975) found that the recognition of the surface level for visual narratives is about 98% with no delay, and after a week it was still remembered better than recognition for written material with no delay. Only after 3 months, the recognition was at chance level (Baggett, 1975). The outcome of this study suggests that the surface level for text fades away sooner compared to surface level memory for a visual discourse.

The second level of processing images for comprehension targets conceptual information (Baggett, 1975). Conceptual information refers to information that is not literally displayed in the image but inferred when a singular picture is integrated into a connected story, e.g., a single panel in a sequence of panels. When processing information on the conceptual information level, the images become meaningful as a whole. This is similar to the text-base level of processing information, in which the information from a discourse gets connected in propositions to form a network of meaning. These are the two levels of image comprehension as discerned by Baggett (1975). The results for the first level of surface information indicate that readers have better memory for the visual discourse as opposed to verbal discourse. Over the years, multiple models have been proposed to explain how image comprehension works in more detail. Two of these are SPECT (Loschky et al., 2019) and the PINS model (Cohn, 2020).

The SPECT and PINS model

The Scene Perception and Event Comprehension Theory (SPECT) aims to explain how people comprehend visual narratives (Loschky et al., 2019) by applying general models of visual cognition to visual narratives. SPECT differentiates between two domains of processing for visual narratives: front-end and back-end cognitive processes. With the front-end processes, information is extracted from an image, and with the back-end processes, this output is used to create a mental model with the information extracted from the image (Magliano, 2020).

Front-end processes are focused on how the eye moves across a visual representation. The front-end processes occur during a single eye fixation (Loschky et al., 2019). The front-end processes determine which information stands out in the image, also called attentional selection. Attentional selection is affected by bottom-up task-driven goals such as searching for specific information in the image and by features facilitating top-down processes sensitive to stimulus saliency. Stimulus saliency attracts attention to certain parts of an image that have a contrast to the rest of the image in terms of color, brightness, color, and size (Loschky et al., 2019).

When the reader has perceived the image with attentional selection, the next step is pulling relevant information from the image, which is information extraction (Cohn, 2019). Information extraction is about the type of information that is extracted during eye fixations. SPECT distinguishes broad information extraction from narrow information extraction (Loschky et al., 2019). Broad information is extracted from the entire scene in the image, which produces semantic information that is the scene gist. This is a general overview of the entire image and includes characters, basic level action, the agent and patient of the action, and general information about the spatiality. Narrow information extraction pulls from one particular entity in the image, which can be a person, animal, or object. The information provides details about this entity such as the shapes, sizes, and colors (Loschky et al., 2019).

The back-end processes construct a coherent mental model with the information that was gathered with the front-end processes. SPECT describes three key processes that are necessary to create this mental model: laying the foundation to create a new mental model, mapping incoming information to the reader’s current mental models, and shifting to create a new mental model (Loschky et al., 2019). The front-end processes happen within one eye fixation, but for the back-end processes multiple eye fixations are needed due to the short time span of a single eye fixation. It takes multiple fixations to extract information that is accurate enough to identify detailed actions that are needed to form a mental model.

The first step, laying the foundation, gathers single pieces of information from the image (Loschky et al., 2019). When a mental model for a new event is created, the reader must lay a foundation. This foundation will likely exist of spatial-temporal information and any agents and actions which are recognized in the first eye fixations. The pieces of information which are extracted during these first fixations, are connected as nodes. These nodes become structures to which succeeding information is connected.

With a foundation set up for the mental model, the reader extends the model by mapping incoming information to the previously connected nodes. The extension is continuous, as new information keeps coming in through the front-end processes. Changes in any event index that are coherent with their current mental model will lead the reader to update or change their model. This means that the model becomes more elaborated with each eye fixation. With this process, the reader needs to unceasingly monitor continuities in the event indices of time, space, entities, causality, and goals (Gernsbacher, 1985).

When discontinuous information is extracted from an image, the mental model will be revised to fit the new information, which is called mapping (Cohn, 2019). When mapping to the current mental model is no longer possible, the incoming information produces a trigger signal. This leads to the continuous activity being parsed into separate events, which is called shifting (Cohn, 2019).

SPECT focuses on how mental models are constructed with perceptual processing while The Parallel Interfacing Narrative-Semantic Model (PINS) explains image processing, with an emphasis on neurocognition. The theories complement each other by each explaining the same process and both adding an extra component to their own theory. The PINS model complements SPECT by adding the narrative structure element of the image into the comprehension model and the SPECT model complements the PINS model by explaining front-end processes.

The PINS model has a focus on two levels of representation: semantic information and narrative structure (Cohn, 2019). There is a narrative structure present in all modalities.

However, as this paper focused on this particular modality of visuals when referring to narrative structure from now on it will be about the narrative structure of visuals, even if it is not specified. The first step is information extraction. The semantic information gets extracted from images when they get comprehended. The PINS model differentiates between three mental processes while comprehending images: information extraction from the image(s), accessing our semantic memory, and creating a mental model. As with SPECT, according to the PINS model, a reader first extracts information from the images, which is called access (Loschky et al., 2019). When a reader accesses information in an image for the first time, they have no prior information yet. Thus, the first access is suggested as the hardest. After that, successive new pieces of information are added to the preceding information that was taken from the image. Access is similar to 'laying a foundation' in SPECT.

The second step of the PINS model is adding the extracted information to the semantic memory (Cohn, 2019). The extracted information becomes mapped into the reader's semantic memory, through the process of semantic access. In other words, semantic access is matching the incoming information to our existing, prior knowledge. As with SPECT, the last step in the PINS model is to incorporate the information from the semantic memory into a growing mental model (Cohn, 2019). This mental model is first situated in the working memory, as long as it is still being constructed (i.e., the reader is still making their way through a sequence of images) and afterward will be stored in the long-term memory, where the mental model of the story is retained into the future (Cohn, 2019).

How we construct mental models is based on different factors, not all of them coming from the readers' side. Authors make deliberate choices in what information is shown in (sequences of) images, and how this information holds up with the rest of the information in the context (Cohn, 2019). This influence of the author is taken into consideration in the PINS model because it can also affect our mental model creation. The way an author portrays information can influence which information readers extract from it, and that is the first step in the PINS model. An author can portray the same piece of information in different ways, and the choice the author makes can lead a reader to extract different information from it. E.g., the author can make choices in how large they want a piece of information to be compared to the rest of the image, or if it gets a muted color or a much brighter color than the rest of the image. These choices are linked to the stimulus saliency, which influences the top-down processes of image comprehension according to SPECT. This means that small choices like this have an impact on information extraction. The design choices by the author lead to a domino effect with at the end of the process, a different mental model.

Within the PINS model, a narrative level of representation goes hand in hand with semantic processing. The narrative level of representation can be explained based on the theory of Visual Narrative Grammar (VNG), which argues that the organization of meaningful information of a visual narrative sequence is similar to how syntactic structure organizes the semantic information of a sentence. Just as grammar gives rules for how to make sentences, the VNG contains rules on how visual sequences can be displayed to package meaning coherently.

The Structure of Visual Narratives

Visual Narrative Grammar (VNG) states that visual narrative sequences are organized comparably to how grammar tells us how to form sentences (Cohn, 2018). In VNG, the most basic sequencing pattern is the canonical narrative schema (Cohn, 2018). In this pattern, the narrative progresses through various stages, which are all parts of the sequence. A sequence usually begins with an Establisher, which is a panel that sets up the scene that is about to start. It introduces the actors, events, and environment of the situation typically in a passive state (Cohn, 2019b). An Establisher is followed by an Initial, which is the anticipation of the event and sets the event in motion (Cohn, 2013). The Initial is followed by the Peak, in which the climax of the sequence taking place. Peak panels typically show the completion of an event or in some cases the interruption of an event, motivating the meaning of the sequence. The last narrative category is the Release, in which the tension is dissolved and the event is wrapped up (Cohn, 2013).

In the canonical schema this particular order of narrative categories is required, but not all the categories are mandatory to be included. There are also narrative categories that can be added into a sequence to make it more complex, which are not part of the canonical schema (Cohn, 2019). One way to do so is with groupings of panels (Cohn, 2019). There can be a top-level arc in which the panels of the sequence follow the canonical narrative schema, and within that overarching schema, groups of panels can establish their own narrative constituent (Cohn, 2019), see Figure 1 for an example. The sequence within a sequence is a center-embedded clause: it can exist alone as a sequence and can be left out of the bigger sequence without disrupting the overarching storyline of the broader sequence.

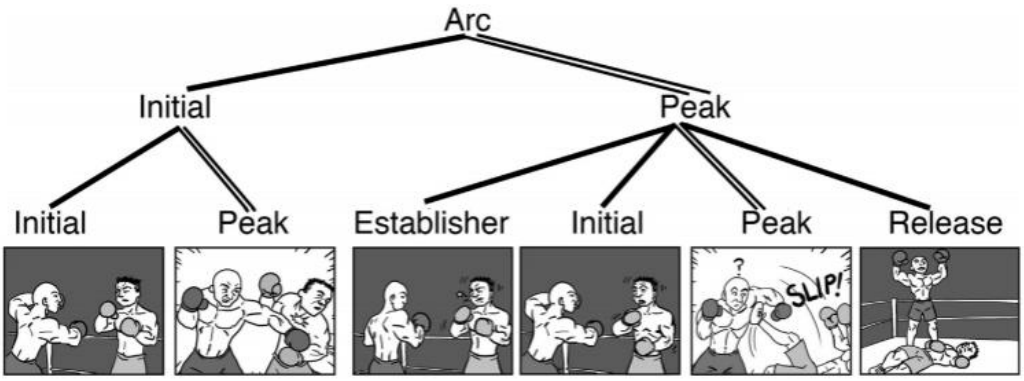


Figure 1. A narrative sequence that exists of two narrative constituents. Adapted from Cohn, 2019. Reprinted with permission.

Besides the canonical narrative schema, there are additional patterns in the VNG that introduce more complex sequences (Cohn, 2018). One of these is the Conjunction. Conjunction repeats narrative categories within a constituent of the same category (Cohn, 2018), which is similar to how syntactic conjunctions repeat grammatical categories, e.g., multiple nouns in a noun phrase (Cohn, 2019b). For example, in Figure 2, panels 2 and 3 show two characters, each in their panel. They play the role of an Establisher by introducing the scene. These panels each draw focus to the individual character, and there are no cues that they belong in the same environment. This must be inferred by the reader by reading the rest of the sequence. This is called the Environmental-Conjunction, or E-Conjunction (Cohn, 2018). Panels 2 and 3 could be substituted for one panel in which you see all three, and the story would still be the same.

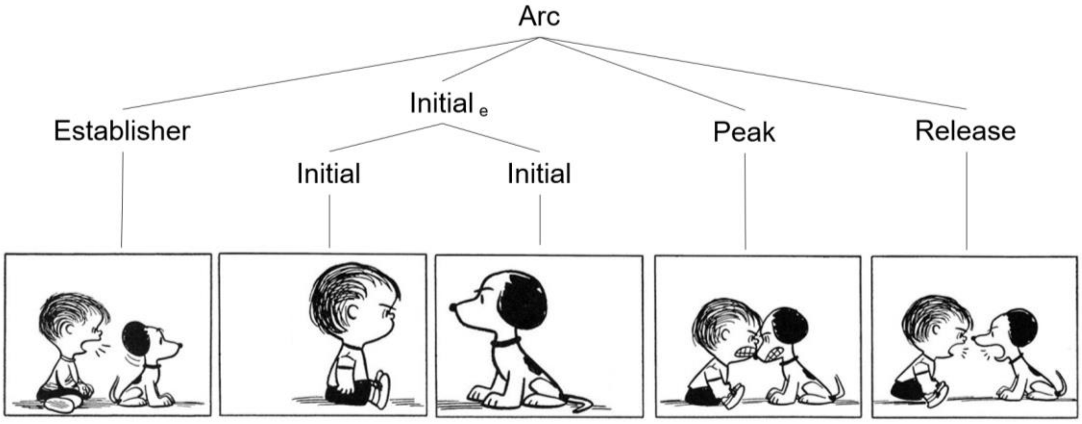


Figure 2. Environmental Conjunction.

Narratively, Conjunction only specifies that a category is repeated within a constituent. Thus, Conjunctions are not only used to break up scenes but can be used for other patterns as well (Cohn, 2018). An Alternation is another pattern that can be used to modify the canonical narrative schema. It exists of repeated pairs of conjoined panels, which form an A-B-A-B pattern that alternates between two characters (Cohn, 2019). Each pairing forms a constituent using Environmental-Conjunction. In Figure 3, you see an Alternation pattern in which the first two panels create the Establisher constituent and the third and fourth panels are the Initial constituent (Cohn, 2019).

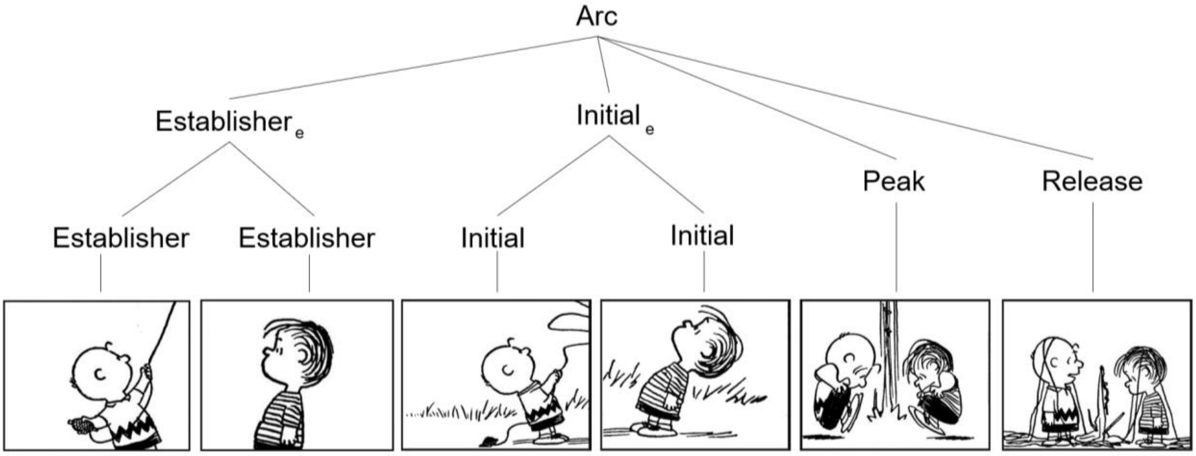


Figure 3. An Alternation pattern.

Another modifier that can be used to increase the complexity of a canonical narrative schema, is the Refiner (Cohn, 2018). A Refiner panel zooms in on information in the preceding panel (the "head" panel relative to the refiner panel). Refiners modify the information by adding extra focus onto a certain part of the preceding panel (Cohn, 2018). E.g., if there is a panel in which a character is about to drop the ice cream that they have in their hand, the refiner zooms in on the ice cream to draw attention to it. To increase the complexity even more, refiners can be placed further from its "head" panel instead of being directly next to its head. This pattern is called Refiner Displacement (Cohn, 2019b).

The Memory of Visual Narratives

With the surface form of text forgotten very fast (van Dijk & Kintsch, 1983) and the research by Baggett (1975) showing that the literal display of images remains longer in our memory, there might be is a difference in how we comprehend information through written or visual discourse. There is a suggested difference in how long surface representation is remembered by readers for written text and visual sequences. However, for visual sequences, it is not clear exactly how well and how long the surface representation stays in our memory.

Currently, there are hints but no evidence that the memory of visual narratives operates the same as the memory of verbal narratives. Baggett (1975) showed that the process of information extraction from images is very similar to how we process information from verbal discourse. However, the study also showed that despite the surface structure gets forgotten quickly for both forms of discourse, the study did find that the surface structure remains longer in memory for visual discourse. Previous studies show similarities, but also some differences. There is conflict, so this study will investigate whether visual discourse gets processed in a similar matter to verbal discourse and what the similarities are.

Another aspect that has not been studied before is the relationship between visual narrative structures and our retainment. A handful of studies have looked at the memory of visual narratives. For example, Gernsbacher (1985) examined how we comprehend information through images, with research on the loss of surface information for individual pictures within picture stories. However, none so far has investigated the relationship between narrative structure and mental models, but rather they have only looked at the surface structure and graphic features of the stimuli. What earlier studies did not account for either was the narrative structure of the whole sequence. Narrative sequences can have different patterns, as explained before. The basic canonical narrative schema can be modified in multiple ways. The modifiers make the canonical narrative schema more complex. The complexity of these various patterns of sequences might be of influence on our image comprehension and our mental models. So far no research has considered the influence of these different patterns on the retainment of the sequences in memory.

The Current Study

The current study seeks evidence for whether the narrative structure of visual narrative sequences is retained in memory and if the complexity of visual narrative patterns is of influence on how well people remember them. Thus, the following Research Questions were formulated:

RQ1: *"Is the narrative structure of visual narrative sequences retained in memory?"*
RQ2: *"Does the complexity of visual narrative patterns influence people's memory performance of sequences?"*

These Research Questions were answered through the following design. Participants were given narrative sequences in two sessions. In session 1, participants were presented with various sequences with different narrative structures (i.e., Basic, Conjunction, Alternation, Refiner

Displacement) and in session 2 participants had to answer if they had seen that sequence in the first session. These patterns with different narrative structures were used because these variations of the canonical narrative schema are a common way of manipulating the surface structure of one original narrative while maintaining the common gist. This way we can see to what degree these variations in surface structure matter. In session 1 participants were asked to rate the comprehensibility of these sequences on a 7-point Likert scale. After a one-week interval, in session 2, participants received the sequences as follows: the exact sequence that saw in session 1, the same sequence with a different narrative structure (e.g. a sequence using the Conjunction pattern in session 1 was provided as an Alternation in session 2) and Semantic Matches were introduced (i.e., a different sequence than in session 1 but semantically relates to the sequences they saw in session 1). The Semantic Match has an associated gist but a different surface structure entirely. In session 2, the participants were asked if they saw the exact same sequence before and how confident they were about their answer on a 7-point Likert scale.

Hypotheses

For Research Question 1, the following hypothesis was formed:

1A. The narrative structure of visual narrative sequences will not be retained in memory, as opposed to the gist, which will remain in memory.

For Research Question 2, two hypotheses were formed:

2A. There will be a difference in perceived difficulty for the narrative patterns.

The second hypothesis for Research Question 2 exists of three competing hypotheses:

2Ba. The more complex the narrative pattern, the harder it is to remember the sequence.

2Bb. The more complex the narrative pattern, the easier it is to remember the sequence.

2Bc. The complexity of the narrative pattern is of no influence on the ability to remember the sequence.

For the second research question, there are competing possibilities for what the outcome could be. This is because there are no comparisons that can be made with previous research. The three competing hypotheses are three possible outcomes, based on discourse research in other domains. Hypothesis 2Ba is based on the classical theories of human memory. According to the classical theories of human memory (Cowan 2001; Cowan 2010) a young adult can remember about 3-5 chunks of meaningful information in their short-term memory. For sequences, this would translate to 3-5 panels. Because more complex sequences have more panels, they cannot be held in our short-term memory and thus not be transferred into our long-term memory.

Hypothesis 2Bb is based on the idea that when a narrative pattern is more complex, participants will take a longer time looking at it to understand it. Having more eye fixations on a sequence means the participant can pull more information from it, and thus remember it better. This hypothesis is based on SPECT, which states that to be able to process information from a picture, eye fixations are needed to extract the information (Loschky et al., 2019).

The final hypothesis, hypothesis 2Bc, is based on the Situation Model view as proposed by Van Dijk and Kintsch (1983). If the comprehension of visual narratives works similar to the comprehension of verbal and written discourse, given that the narrative patterns refer to the surface structure, the surface structure might not be retained in memory at all and thus be of no influence on how well people remember sequences.

Method

Participants

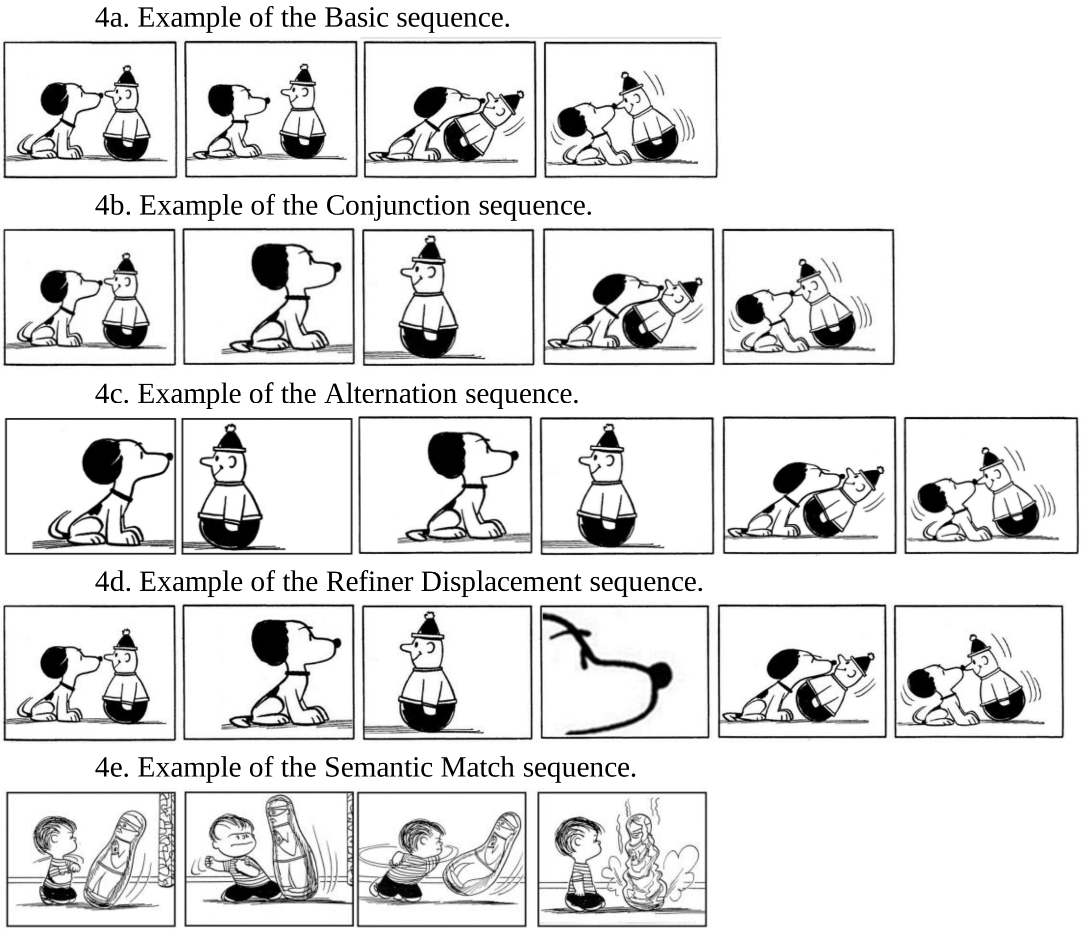
Participants came from the participant pool of the faculty of Humanities and Digital Sciences at Tilburg University. Of the 74 participants, 57 filled out both studies, and their data was used for the data analysis. Of these participants, 17 were male and 40 were female. The average age of the participants was 20.7 (*SD* = 2.87).

To be eligible to participate in this experiment, the participants had to have experience with reading comics. To assess their proficiency in comics, participants filled out a questionnaire designed to calculate their "comic reading fluency" (Visual Language Fluency Index or VLFI). In this questionnaire, participants were asked to rate the frequency of reading various types of visual narratives and drawing comics, both currently and while growing up. These ratings were measured using a 7-point Likert scale, and the questionnaire also gauged their self-assessed "expertise" at reading and drawing comics along a five-point scale. An idealized average along this metric would be a score of 12, with low being below 7 and high above 20. Participants' fluency was a high average, with a mean score of 12.94 (*SD* = 6.40).

Stimuli

For the stimuli, 36 comics were selected from a corpus of sequences created using panels from The Complete Peanuts volumes by Charles Schulz (1950-1974). None of these sequences had text, or text had been removed. For each of the 36 sequences, 4 versions were created with the four narrative patterns: a Basic sequence, a Conjunction sequence, an Alternation sequence, and a Refiner Displacement sequence. For session 2, all sequences from session 1 were reused and a Semantic Match sequence was added for each of the 36 sequences.

The Basic sequence consisted of four panels, one for the Establisher, Initial, Peak, and Release (see Figure 4a). For the Conjunction sequence, the second panel (the narrative Initial) was split into two panels, each showing a single character (Character "A" and Character "B") (as in Figure 4b). For the Alternation sequence, both the first and second panels were split into two panels so that they all had only one character in an A-B-A-B pattern (see Figure 4c). And for the Refiner Displacement, a fourth panel was added which zoomed in on an important feature of the "A" character ("a") in the second panel to create an A-B-a pattern (see Figure 4d). The Semantic Match for each sequence was a different yet semantically related sequence (see Figure 4e). According to the Event Indexing Model (Zwaan, Langston & Graesser, 1995) readers update their mental models with several indices when receiving new information. The five types of indices that get updated are temporality, spatiality, protagonist, causality, and intentionality. Thus, for the Semantic Matches, sequences differed on at least one of these indices from the semantically related sequence it was matched to. For example, if one of the sequences was about Lucy playing baseball, the Semantic Match was a sequence about baseball but it was slightly different, e.g. with Snoopy playing instead of Lucy. This would be a difference in protagonist according to the Event Indexing Model.



4a. Example of the Basic sequence.

4b. Example of the Conjunction sequence.

4c. Example of the Alternation sequence.

4d. Example of the Refiner Displacement sequence.

4e. Example of the Semantic Match sequence.

Figure 4. Example of one sequence in five narrative patterns used in the study.

In session 1, each participant saw 36 sequences, 9 of each narrative pattern. There were 4 lists, which were randomized so that for each of the 36 sequences, each one was shown in all 4 narrative patterns across the lists with a Latin Square Design. The Semantic Matches were not used yet in session 1. Sequences were also randomized across participants so within each list, participants saw the same sequences in unique orders.

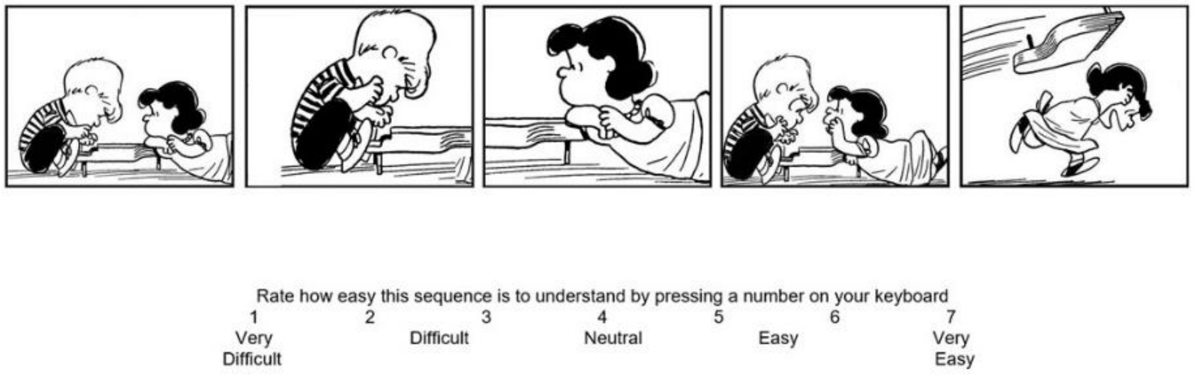
In session 2, the same sequences were used, again in the 4 narrative patterns along with the Semantic Matches. In this session, participants either saw the same sequence in the same narrative pattern they saw in session 1 (i.e., Same condition), the same sequence in a different narrative pattern than they saw in session 1 (i.e., Different condition), or the Semantic Match of the sequence they saw in session 1 (i.e., Semantic Match condition). Five lists were created to counterbalance the narrative patterns, again with a Latin Square Design. The lists were designed so that each list had an equal number of Same – Different – Semantic Matches and within that, each had an equal number of patterns.

Procedure

The questionnaire through which the experiment was taken was available in the participant pool of the Humanities and Digital Sciences faculty of Tilburg University. On the website of the participant pool, participants could sign up for both session 1 and session 2, in exchange for 0.5 credit. After they signed up, they could start the first session through the Qualtrics link. Participants had to come up with a unique code in session 1, which they had to remember for session 2. Exactly one week after they completed session 1, they were notified that they could participate in session 2. Credits were granted when participants completed both sessions.

Session 1

In the Qualtrics survey, participants were given general information about the experiment and asked to give consent to participate. Then they had to answer demographic questions and questions about their experience with reading comics. After this, the experiment started. In both sessions, the participants received 36 sequences. In session 1, they saw sequences on the screen with the instruction 'rate how easy this sequence is to understand by pressing a number on your keyboard' and a 7-point Likert scale (1 = very difficult, 7 = very easy) under the sequence. See Figure 5 for an example. They then had to press the corresponding number on their keyboard to proceed to the next sequence. Between each sequence was a rest screen so that the participants could see how many sequences were left in the session.



Rate how easy this sequence is to understand by pressing a number on your keyboard

1	2	3	4	5	6	7
Very Difficult		Difficult	Neutral	Easy		Very Easy

Figure 5. A screen from session 1.

Session 2

In the Qualtrics survey for session 2, participants again started on a screen with general information and were asked to give consent. Then they proceeded to the experiment, as the demographic information had already been collected in session 1. In session 2, the participants again saw 36 sequences. This time, each sequence was provided on the screen twice. The first time the sequence was on the screen, under it was the question ‘Did you see this sequence in the previous session?’ which they had to answer with yes or no. Then they saw the same sequence again, with the question ‘How confident are you that you saw this sequence in the previous session?’ and a 7-point Likert scale. After answering both questions, there was a rest screen and then the next sequence appeared.

Data Analysis

Not all the data of participants were used in the analysis. For the analysis, response times that were too slow or too fast were removed. These were the response times that differed more than 2 standard deviations from the average response time of all participants. Due to a technical error in Qualtrics not all the responses in session 2 were recorded. For each participant, about 1 to 5 trials of the 36 were lost. Because only a small portion of the data was lost, this did not form an issue for the data analyses of session 2. The data analysis for this experiment was done in Jasp (Jasp Team, 2020).

To investigate various relationships between variables, the following tests were used. For session 1, to investigate the differences in comprehensibility between the sequences, a Repeated Measures ANOVA was used with independent variable Pattern and dependent variable Comprehension Rating. To investigate whether there was a difference in response time for the different sequences, another Repeated Measures ANOVA was used with independent variable Pattern and dependent variable response time. A correlation analysis was performed to investigate whether the VLFI score correlated with the comprehension rating.

For session 2, a Repeated Measures ANOVA was used to investigate if participants remembered whether they had seen the sequences in session 1 with the Condition as independent variable and recognition answer as dependent variable. The confidence of the participants for their choice of answer was measured with a Repeated Measures ANOVA, with Condition as independent variable and Subjective Confidence as dependent variable. Then a One-Way ANOVA was performed to look into whether the recognition answers were actually correct, with Condition as independent variable and accuracy score as dependent variable. Then it was investigated whether the different levels of perceived difficulty influenced how well people remember them with a Repeated Measures ANOVA, with accuracy scores for the patterns as dependent variables and the condition (i.e., same or different) as independent variable. The relation between response times for session 1 and session 2 was compared with a correlation. And lastly a multiple regression analysis was conducted to investigate whether there was a relationship between VLFI score, comprehension rating and session 1 response time as predictors and the recognition answer in session 2 as outcome.

Results

Session 1

Comprehension Rating

In session 1, participants were asked to rate how easy they thought the sequences were to understand on the 7-point Likert scale (1 = very difficult, 7 = very easy), to investigate if there is a difference in the perceived level of difficulty between the different patterns. Participants rated the sequences overall as fairly easy to understand, with a mean rating of 5.32 (*SD* = 0.93). For the mean rating per narrative pattern, see Table 1. As predicted, participants rated the Basic sequence as easiest to understand and the Refiner Displacement sequence as most difficult, with Conjunction and Alternation sequences in between.

Pattern	Mean	SD
Basic	5.45	1.13
Conjunction	5.37	0.94
Alternation	5.16	1.09
Refiner Displacement	4.98	1.01

Table 1. The mean comprehension ratings per narrative pattern.

A One-Sample T-Test was performed for each condition, to see whether the comprehension ratings per pattern differed significantly from the midpoint answer on the 7-point Likert Scale which corresponds to the score 4. Scores above the midpoint indicate that the sequence is comprehensible, scores below the midpoint indicate that the sequence is not comprehensible. All four patterns were rated significantly higher than the chance level (for all patterns: $t(58) > 9.71$, $p < .001$), showing that participants thought all the patterns were understandable.

Comprehensibility ratings were compared across patterns with a Repeated Measures ANOVA. There was a main effect of Pattern for Comprehension Rating, $F(3, 168) = 9.78$, $p < .001$, $\eta^2 = .15$. A Bonferroni pairwise post hoc analysis showed that the Basic pattern (*M* = 5.45, *SD* = 1.13) was significantly easier to comprehend than the Refiner Displacement pattern (*M* = 4.98, *SD* = 1.01), $p < .001$. The Basic pattern was also significantly easier to understand than the Alternation pattern (*M* = 5.16, *SD* = 1.09), $p = .019$. The Conjunction pattern (*M* = 5.37, *SD* = 0.94) was also significantly easier to understand than the Refiner Displacement pattern, $p < .001$, see Figure 6. No other significant differences were observed between patterns, all $p > .219$.

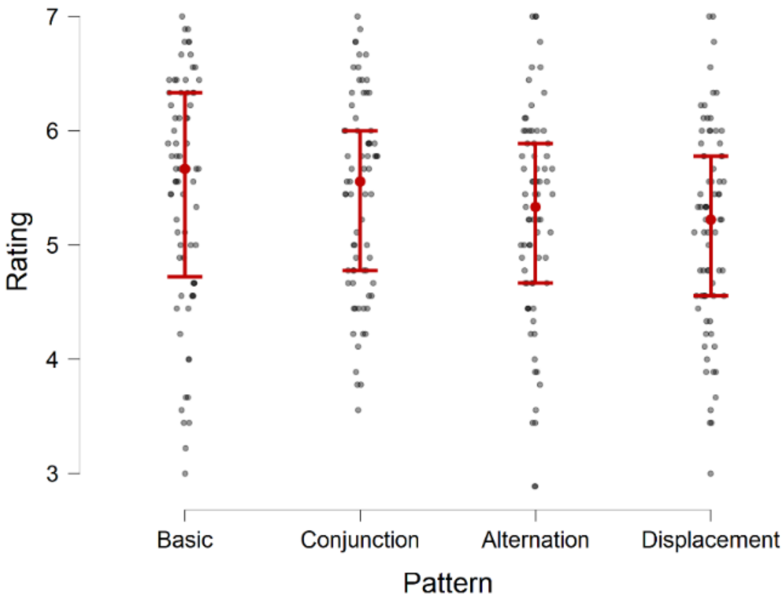


Figure 6. The mean comprehension ratings per pattern. Note: scale zoomed in on the y-axis due to graphing program.

These results show that the Basic pattern is perceived to be the easiest to understand, and the Refiner Displacement is harder to understand than the other narrative patterns. The Conjunction and Alternation sequences are placed in the middle, with the Conjunction being slightly easier to understand than the Alternation patterns.

Response time

The time participants spent to rate the sequences (i.e., response time) was also compared across patterns with a Repeated Measures ANOVA. The response time was the time a participant spent on one trial, which comprises of looking at the sequence and then rating it. The mean response time was 6818.70 milliseconds (*SD* = 2579.89). The mean response time per narrative pattern can be seen in Table 2.

Pattern	Mean	SD
Basic	6490.94	2612.10
Conjunction	6872.92	2473.63
Alternation	6907.06	2333.74
Refiner Displacement	7259.93	2456.45

Table 2. The mean response time per pattern.

There was a main effect of Pattern for Response Time, *F*(3, 168) = 4.77, *p* = .003, η^2 = .08. A Bonferroni post hoc test showed that the participants had significantly faster response times to the Basic sequences (*M* = 6490.94, *SD* = 2612.10) than the Refiner Displacement sequences (*M* = 7259.93, *SD* = 2456.45), *p* = .001. Between the other conditions, there were no significant differences, all *p* > .255.

Effect of Comic Reading Proficiency (VLFI)

Participants' proficiency in reading comics was measured with the Visual Language Fluency Index (VLFI). From their answers, a VLFI score was calculated, which showed how proficient participants were at reading comics. The higher the VLFI score, the more proficient someone is at reading comics. The proficiency of reading comics might be of influence on how comprehensible participants think the sequences are. To test if the proficiency of reading comics (VLFI score) correlates with the comprehension rating, a correlation analysis was performed. The analysis showed no correlation between participants' VLFI scores and comprehension ratings (*r* = .04, *p* = .284), nor with response times (*r* = -.02, *p* = .630).

Session 2

Recognition

To test how well participants remembered if the sequence they saw in session 2 was the same as in session 1 (i.e., recognition), a Repeated Measures ANOVA was used. During the trials, participants had to answer with yes (1) or no (0) to the recognition question. The recognition scores are regardless of accuracy, they just measure the 'Yes' responses, regardless of whether they are actually correct. There was a main effect of Condition for Recognition, *F*(2, 112) = 95.71, *p* < .001, η^2 = .63. A Bonferroni post hoc test showed that participants remembered the Same condition (M = 0.61, SD = 0.18) significantly better than sequences in the Semantic Match condition (M = 0.24, SD = 0.16), *p* < .001. The Different condition (M = 0.60, SD = 0.21) was also remembered significantly better than the Semantic Match condition, *p* < .001. There was no significant difference between the Different and Same conditions, *p* = .958. These results show that the Semantic Match condition got significantly more 'no' ratings, meaning that participants remembered they did not see it in session 1, see Figure 8.

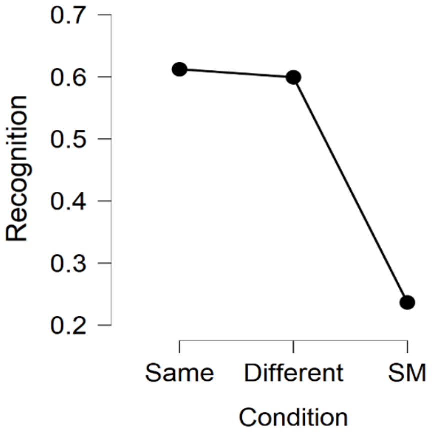


Figure 8. The average of Recognition per condition.

The time it took participants to answer the recognition question (i.e., response time) was also measured per condition with a Repeated Measures ANOVA. There was no main effect between Conditions for Response Time, *F*(2, 112) = 0.87, *p* = .422, η^2 = .02. See Table 3 for the response time per condition. This result shows that the participants were equally quick for answering the recognition question across the different conditions. For each condition, a participant needed about the same amount of time to decide whether they had seen the sequence before.

Condition	Mean	SD
Same	4433.23	1393.78
Different	4571.30	1433.44
Semantic Match	4571.87	1395.56

Table 3. The mean response time in milliseconds per condition.

Subjective Confidence

After participants answered the recognition question, they were also asked to rate how confident they were about their choice on a 7-point Likert scale (1 = very sure, 7 = not sure). This Confidence Rating was the second question per trial in session 2, with which participants could specify how sure they were about their recognition. This subjective confidence rating was compared across patterns with a Repeated Measures ANOVA. There was a main effect of Condition for Subjective Confidence, *F*(2, 112) = 2.68, *p* = .073, η^2 = .05. A Bonferroni post hoc test showed that participants were more confident about whether they had seen the sequence in session 1 for Different condition (*M* = 3.16, *SD* = 0.93) than the Semantic Match condition (*M* = 3.39, *SD* = 0.96), *p* = .042. There was no significant difference in how confident the participants were about their choice for the Same condition (*M* = 3.24, *SD* = 0.95) compared to the Different condition or Semantic Match condition, all *p* > .444. These results show that participants were most confident about their choice of answer for recognition in the Different conditions. They were in doubt the most about their answers in the Semantic Match condition, see Figure 9.

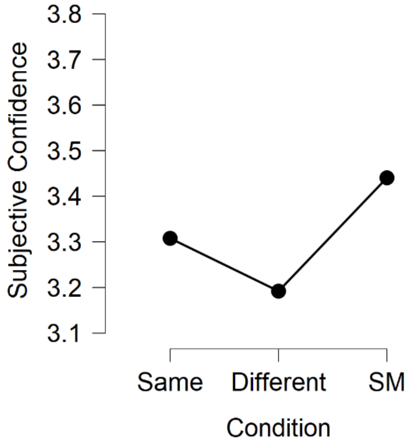


Figure 9. The subjective confidence answers about recognition. Note: scale zoomed in on the y-axis due to graphing program.

The response time for the subjective confidence was considered as well. The response time for the subjective confidence question was analyzed with a Repeated Measures ANOVA. There was a main effect of Subjective Confidence on Response Time, *F*(2, 112) = 3.86, *p* = .024, η^2 = .06. A Bonferroni post hoc test showed that participants were significantly slower at answering the subjective confidence question for the Semantic Match condition (*M* = 1870.00 *SD* = 653.76) than the Same condition (*M* = 1661.78, *SD* = 670.44), *p* = .043. The response times did not differ significantly from the Different condition (*M* =1675.98, *SD* = 578.67), all *p* > .067. See Figure 10.

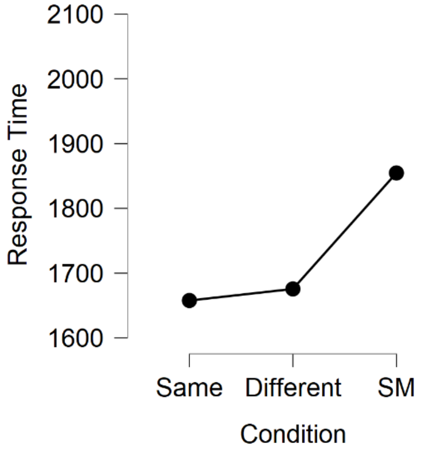


Figure 10. The mean response times per condition.

Accuracy of Answers

An analysis was performed on recognition to see how often participants answered ‘Yes, I did see this sequence in the previous session’, regardless of whether participants saw it. Those results did not show how often their answers were actually correct. A new variable was created to see how often the results were correct. Accuracy of answer was calculated by taking the mean score of correct answers per participant per condition. ‘Yes’-answers were correct for the Same condition and ‘No’-answers were correct for the Different and Semantic Match conditions. The range of the accuracy was 0 to 1, with 0 being no correct answers for that condition and 1 being all correct answers for that condition. A One-Way ANOVA analysis was conducted to see if there is an influence of condition on recognition. There was a main effect of Accuracy, *F*(2, 510) = 49.70, *p* < .001, see Figure 12. A Bonferroni post hoc test showed that the memory performance was significantly higher for the Semantic Match condition (*M* = 0.78, *SD* = 0.16)than the Same condition (*M* = 0.61, *SD* = 0.30), *p* < .001. Participants were significantly more accurate in their answers for the Same condition than the Different condition (*M* = 0.40, *SD* = 0.33), *p* < .001.

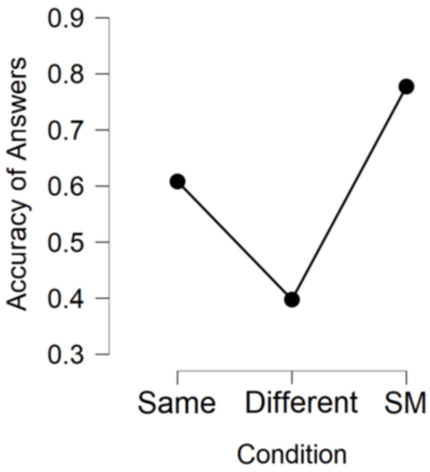


Figure 12. Accuracy of answers across sessions.

Narrative Patterns

As established in session 1, the different patterns have different levels of perceived difficulty. To see whether the difficulty of these patterns influenced how well participants remember them, a Repeated Measures ANOVA was conducted with accuracy scores for the different patterns, in both the Same and Different conditions. In the Same condition, participants saw the same sequence they had seen in session 1. There was no main effect of Pattern on Accuracy in the Same condition, *F*(3, 168) = 0.97, *p* = .410, η^2 = .02, nor a main effect of Pattern on Accuracy in the Different condition, $F(3, 168) = 1.94$, $p = .125$, $n^2 = .03$. The mean accuracy scores can be found in Table 4.

	Same		Different	
Pattern	Mean	SD	Mean	SD
Basic	0.56	0.31	0.35	0.32
Conjunction	0.62	0.29	0.42	0.32
Alternation	0.61	0.28	0.36	0.32
Refiner Displacement	0.65	0.31	0.46	0.34

Table 4. The mean accuracy rating per pattern in the Same condition, with 0 being incorrect and 1 being correct.

Response times

Response times in session 1 and session 2 were compared. To see if participants with longer response times in session 1 have a shorter response time in session 2, a correlation was conducted. For session 1, this is the response time it took to answer the question of comprehension rating, and for session 2 this is the response time for the first question about recognition. Based on the results of the study, participants who took a longer time responding in session 1, also took a longer time responding in session 2, $r = .28$, $p < .001$. See Figure 14. This result shows that participants who took their time to answer the question in session 1, were also slower to answer the recognition question in session 2.

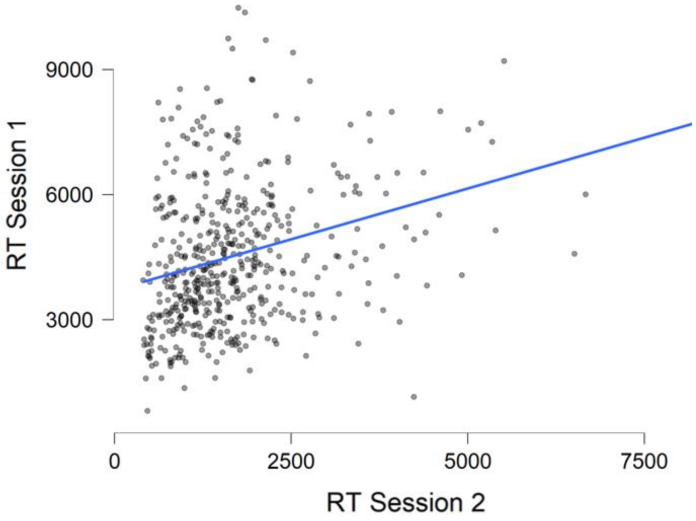


Figure 14. The correlation between response time in session 1 and response time in session 2.

Effects on Recognition

To investigate if there is a relationship between VLFI score, comprehension rating, and session 1 response time as predictors and the recognition answer in session 2 as the outcome, a multiple regression analysis was performed. The results show a significant effect on Recognition, $F(3, 508) = 26.34$, $p < .001$, with R^2= 0.130, suggesting that 13.0% of the variance in recognition can be accounted for by the three predictors collectively.

Comprehension ratings and response times of session 1 both predict recognition in session 2. Response time predicts recognition in a positive way ($\beta = .157$, $t = 3.79$, $p < .001$), and comprehension rating predicts recognition in a negative way ($\beta = -.326$, $t = -7.90$, $p < .001$). The regression analysis showed that the VLFI score was no predictor of recognition ($\beta = .054$, $t = -1.31$, $p = .192$).

Discussion

This study examined the idea of whether the narrative structure of visual narrative sequences is retained in memory, and if the complexity of visual narrative patterns is of influence on this retainment. The main finding of this study is that the narrative structure of visual narratives is not retained in memory as opposed to the gist, which is retained. The complexity of the different visual narrative patterns is of no influence on the retainment of this memory.

Memory of the sequences

The primary question of this study was whether participants would remember the surface structure of a visual narrative sequence. The recognition of participants, which is to say, could they remember if they had seen a specific sequence in the first session, was measured in the second session. Participants were less accurate in recalling whether they had seen the sequence before for both the Same and Different conditions. They were most often correct in their answer for the Semantic Match condition. That the Semantic Match condition is remembered most accurately might indicate that participants remember gist more than the narrative pattern and have a hard time differentiating between sequences that only differ in narrative structure, as the Semantic Match conditions were selected to look similar to their matched sequence but had a slight difference in meaning.

This result is in line with hypothesis 1A. The results suggest that the narrative structure of visual sequences is not retained in memory, as opposed to the gist. This conclusion can be taken from the recognition results because participants do remember that they did not see the Semantic Match condition before, which has a different gist than the sequence they saw in session 1. Participants were worse at remembering if they had seen either the Same and Different conditions in session 1, suggesting that participants could not differentiate between sequences when the only difference was the narrative structure. The gist of the story remained the same in these conditions, and this was remembered. The differentiating part, the pattern of the sequence, which they needed to remember to say if they saw the sequence before, was forgotten. This result is in line with the discourse research of Van Dijk & Kintsch (1983) and Zwaan and Radvansky (1998) who both argue that what we store in our mental models is the gist and that the surface structure is forgotten.

Another interesting observation was that of the correlation between the response times of session 1 and session 2. There was a positive correlation between response time in session 1 and session 2, meaning that if participants took a longer time to rate the sequence in session 1, they also took a longer time to answer the recognition question in session 2. This result is not in line with the proposed idea that when participants took a long time looking at the sequence in session 1, they should remember it better and consequently be able to make a faster decision in session 2 for the recognition question. A reason for this might be that some people are slower in general with reading and answering questions and want to take their time before answering.

Further insight into participants' thinking about their recognition processes was gained from analyzing their subjective confidence ratings. Here, they were most confident about their recognition for Different conditions, despite the relative inaccuracy of those ratings. This was closely followed by the Same condition. Participants were less sure about their answer for the Semantic Match condition. Although participants were most accurate for the Semantic Match sequences that they had not seen in session 1, they were least confident about those sequences. This might indicate that because the gist is retained, they could differentiate that they had not seen the Semantic Matches in session 1. However, because the gist was similar to the gist of the originally given sequences, that might lead to lower confidence about their memory. On the other hand, when participants are presented with the same sequences with a different narrative structure, they had a hard time differentiating those from the ones with the original narrative structure. But they were confident about their wrong answers, which shows that the structure was not stored in memory. Together, these findings also indicate that narrative structure was not retained in memory, but the gist retained.

The recognition results were enhanced by the response times. The response time for the confidence question was faster in both the Same and Different conditions and the slowest for the Semantic Match condition. That the response time was significantly slower for the Semantic Match conditions shows that participants indeed were most in doubt about their answer and took longer to decide. That there was barely any difference in the response time for the Same and Different conditions shows that participants had a hard time differentiating between those two conditions. These results suggest that participants do not remember the difference between the Same and Different conditions, since those had the same response times. This implies that participants could not differentiate between the difference in surface structure, which was the only difference between the Same and Different conditions. These results again suggest that the surface structure does not get retained in memory.

Compared to other forms of discourse, this shows similarities between visual narratives and other types of discourse. Just as with verbal discourse, the surface structure of visual narratives gets forgotten quickly. This seems to be a domain-general process with the surface structure disappearing in a similar manner. That the narrative structure is not persisting, raises the question of what its purpose is. As with spoken language, the grammar and form of the discourse are forgotten very fast because it serves as a tool to carry the meaning. The Visual Narrative Grammar, just like grammar for verbal discourse, seems to be a tool for carrying information. As soon as the meaning is stored by the recipient, the means of transferring is no longer important and forgotten.

Narrative Structures

The second question was whether narrative patterns differed in how difficult they were to understand and if this complexity was of influence on people's memory performance of the sequences. Participants rated the Basic sequences as easiest to understand and Refiner Displacement as most difficult. The ratings for the Conjunction and Alternation fell in the middle, with Conjunction being perceived as slightly easier than the Alternation sequences. The response time in session 1 further informs the rating. The response times were the fastest for the Basic sequence and slowest for the Refiner Displacement sequence. The response times for Conjunction and Alternation were in between the other two and did not differ much from each other. These results are in line with the results of the comprehension ratings. The more difficult the pattern was rated, the longer it took participants to rate it. These results suggest that it is possible to make more variations of the canonical narrative schema (i.e., the Basic sequence) with the VNG that are more complex to understand. This is in line with the results of Cohn (2019) and Cohn (2018) which both suggest that the VNG introduces more comprehensive sequences other than the canonical narrative schema.

The results also imply that the more comprehensive sequences are harder to understand than the canonical narrative schema. The Refiner Displacement is rated as most difficult, and the Conjunction and Alternation are both slightly less difficult. The Refiner Displacement and Alternation patterns are the two most difficult sequences and both 6 panels long, but the Refiner Displacement is perceived as more difficult than the Alternation sequences. This means that the panel length is not the factor that affects the difference in perceived difficulty. The thing that makes the Refiner Displacement the most difficult might be that understanding the sequences requires the participants to understand two levels of modifiers: conjunction and refiner. The other sequences have no modifiers (i.e., Basic sequence) or only one modifier (i.e., Conjunction and Alternation sequences). However, this is only a suggestion and cannot be said with certainty since it was not studied in this experiment.

These results confirm hypothesis 2A, the different patterns of sequences that can be made with the VNG have different levels of difficulty. As more complex sentences are created with grammar, the VNG can be used to create patterns with different degrees of comprehension difficulty (Cohn, 2018). This result is in line with the VNG, which proposes that drawings use similar structural principles as language and thus can be made more difficult to comprehend with the use of surface structure.

The response times were informative for the session 1 results. However, it is a possibility that the response times were influenced by the number of panels per pattern. The Basic sequence exists of 4 panels, the Conjunction sequence is 5 panels long and both the Alternation and Refiner Displacement are 6 panels long. The longer response time for sequences with more panels might be caused by participants needing more time to look at all the panels. A correlation found this was indeed the case, and the response time was longer the more panels a sequence had (*r* = .18, *p* = .013), see Figure 16. This suggests that in session 1, the response times were affected by the number of panels instead of, or as well as by the perceived difficulty of the narrative patterns.

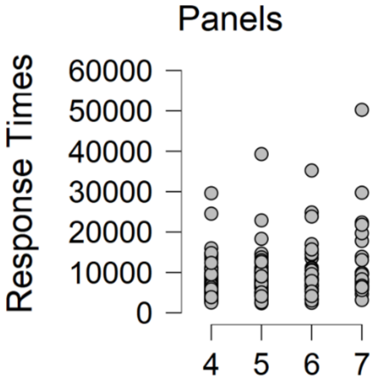


Figure 16. The average response time per number of panels in a sequence (4 to 6 panels). The sequence with 7 panels was an exception. One Basic sequence had 5 panels instead of 4, so each pattern had one more panel.

Future studies might want to control for the number of panels, as it is an influence on the response times. Length could be controlled through different manipulations. The first option is prolonging the shortest sequence with an Orienter, which is a panel providing superordinate information about the sequence (Cohn, 2013), usually placed at the beginning or end of a sequence. This would not change the gist of the sequence but add another panel with general information. With an Orienter the Basic sequence could be 5 panels long instead of 4, but then there would still be a difference between the patterns as they would have 5 or 6 panels.

Another option to control for length would be to show a sequence panel by panel and letting the participant be in control of clicking to the next panel. Then the response time for the Peak can be measured on its own, which is the main event. Since the differentiation between the different patterns happens in the Establisher or Initial, these modifications should influence how fast participants make the narrative connection between the preceding panels and the peak, thus understanding the main event of the story. The response time of the Peak can be used to see how fast participants can connect the preceding panels to the most important part of the story.

Thus far, it has been established that the narrative structure of visual sequences is not retained in memory. However, is it possible that, despite this, the complexity of narrative patterns influenced how well participants remembered the sequences? If easier patterns are more comprehensible, it would be in line with the Situation Model view (Van Dijk and Kintsch, 1983) which suggests that we remember the gist of what we receive. For easier patterns, participants receive less information about the surface structure because these have a small number of panels and these are thus easier to store in memory (hypothesis 2Ba). Or if complex patterns are easier to comprehend and remember because they require more eye-fixations and thus readers can make a more extensive mental model, it would be consistent with SPECT which posits that with every eye-fixation, we pull more information from an image. This means that if there are complex patterns with more panels, people have to use more eye-fixations and thus get more chances to store and memorize information about the sequence (Loschky et al., 2019) (hypothesis 2Bb). However, it has been established that the narrative structure of visual narrative sequences is not retained in memory, so there can not be an effect of narrative patterns on a non-existent memory.

The results of session 2 showed that there was no difference across patterns for the correct answers. The accuracy of answer ratings barely differed per pattern, so there was no difference in how well participants remembered the different narrative structures and the narrative structure was of no influence on how well participants remember them. This result is further supported by the analysis across the Same and Different conditions. For both of these conditions, the accuracy of answer scores were about the same for each of the patterns. This indicates that the complexity of surface narrative pattern does not affect the memory of the meaning. The PINS model focuses on two levels of representation: semantic information and narrative structure (Cohn, 2019). These results show a broader separation between narrative structure and semantics. Since the narrative patterns do not matter for the memorization of meaning, this implies a difference between those two factors. The results thus support the overall idea of the PINS model.

As expected by the outcome of Research Question 1, the results showed no significant difference in how well people remember the different types of patterns accurately, in both the

Same and Different conditions. The analysis suggests that the narrative structure of visual narrative sequences does not get remembered. This means it is reasonable to believe there would be no difference in how well people remember these narrative structures with different levels of difficulty because they do not get retained in memory at all. With these results, hypothesis 2Bc can be confirmed. The complexity of the narrative pattern is of no influence on the ability to remember the sequence. This hypothesis was motivated by the Situation Model view (Van Dijk & Kintsch, 1983). The degree of difficulty of the narrative sequence is of no influence on how well people remember it because the surface structure does not get retained.

Despite memory not differing for narrative patterns, an interaction response to the recognition question did differ between patterns for different levels of VLFI scores. The VLFI scores of participants were calculated to see if fluency in reading comics would be of any effect on their comprehension rating of the different sequences and the ability to differentiate between the different sequences. For both the Basic and the Refiner Displacement sequences there was a relation between VLFI and recognition. For both sequences, participants with a higher VLFI score were more inclined to answer 'yes' for the recognition question. For the Conjunction the opposite happened. The higher the VLFI score, the less likely participants were to answer 'yes' for the recognition question. There was no effect of VLFI on the recognition answer for Alternation sequences.

Basic sequences and Refiner Displacements are on the far ends of the comprehensibility scale, with Basic sequencing being the easiest to comprehend and Refiner Displacement the most difficult to comprehend. These results might be the effect of this. The comprehensibility of Conjunction and Alternation sequences was rated in the middle and for those, there is no effect. These results imply that the most simple sequences might be easy to remember due to their simplicity, which makes them easy to comprehend and remember and the more difficult sequences might be easy to remember because they are so difficult, meaning a reader spends more time looking at it and thus can create a better mental model for the gist.

These results can be linked to previous research on proficiency in reading comics. According to Cohn (2020), people with a higher proficiency might store patterns in memory more explicitly. This would mean that people with a higher VLFI score store patterns more explicitly, and people with a lower VLFI score store fewer patterns. This implies that the reason people with a higher VLFI score were more inclined to answer 'yes' for the recognition question for Basic and Refiner Displacement sequences and less inclined to answer 'yes' for the Conjunction, is because they store the patterns more explicitly and thus the patterns that stand out (i.e. Basic and Refiner Displacement) due to their length and comprehensibility, are easier to remember.

Limitations and Future Directions

Overall, we looked at whether the surface structure of visual sequences with a narrative structure was retained. This research examined two primary research questions. Would the narrative structure of visual narrative sequences be retained in memory like the gist and if so, would the difference in perceived difficulty of the narrative patterns be of influence on how well we remember these narrative structures? The results of this research are consistent with what is found in general discourse literature such as the Situation Model view (Van Dijk & Kintsch, 1983; Gernsbacher, 1985; Zwaan & Radvansky, 1998) meaning that narrative structure fades from memory in the same way that is hypothesized for verbal and written discourse. Thus the process of forming a mental model of the gist seems to be a domain-general process and is tied in with general principles of discourse and how we retain it in memory. This also means that the second research question was answered, as the results of the first research question already showed that the surface structure was not retained. The patterns were of no influence since the memory was non-existent.

This approach shows the benefit of having an explicit theory of the structure of narratives that can be examined across different structures rather than a general notion of surface structure. The similarities between visual discourse and other discourse have been addressed, but any potential differences between visual narratives and other forms of discourse can not be addressed with the results of this study as they were not the focus. To investigate possible differences further research is needed. Further studies could better assess the direct differences between modalities.

Another question that remains unanswered after this research is whether the same goes for visual sequences without narrative structures, i.e. sequences of images without any related meaning. Nothing can be said about whether the narrative structure has an advantage over random, scrambled sequences of panels with regard to remembering the surface structure. Based on other discourse literature, the expectation is that it is harder to create mental models of incoherent information than coherent information (Gernsbacher, 1985; Cohn et al., 2012). However, little information is known about what happens to the memory for the surface structure of scrambled sequences in comparison to coherent narratives. This question remains unanswered and open for further research.

References

Baggett, P. (1975). Memory for explicit and implicit information in picture stories. *Journal of Verbal Learning and Verbal Behavior, 14*(5), 538-548.

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

Cohn, N. (2013). *The Visual Language of Comics: Introduction to the Structure and Cognition of Sequential Images*. London, UK: Bloomsbury.

Cohn, N. (2018). In defense of a “grammar” in the visual language of comics. *Journal of Pragmatics, 127,* 1-19.

Cohn, N. (2019). Visual narratives and the mind: Comprehension, cognition, and learning. In *Psychology of learning and motivation* (Vol. 70, pp. 97-127). Academic Press.

Cohn, N. (2019b). Structural complexity in visual narratives: Theory, brains, and cross-cultural diversity. *Narrative complexity and media: Experiential and cognitive interfaces*, 174-199.

Cohn, N. (2020). Visual narrative comprehension: Universal or not? *Psychonomic bulletin & review*, *27*(2), 266-285.

Cohn, N., & Magliano, J. P. (2020b). Editors’ introduction and review: Visual narrative research: An emerging field in cognitive science. *Topics in Cognitive Science*, *12*(1), 197-223.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and brain sciences, 24*(1), 87-114.

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current directions in psychological science*, *19*(1), 51-57.

Gernsbacher, M. A. (1985). Surface information loss in comprehension. *Cognitive psychology*, *17*(3), 324-363.

Hemforth, B., & Konieczny, L. (2006). Language processing: construction of mental models or more?. In *Advances in Psychology* (Vol. 138, pp. 189-204). North-Holland.

JASP Team (2020). JASP (Version 0.14.1)[Computer software].

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological review, 85*(5), 363.

Loschky, L. C., Larson, A. M., Smith, T. J., & Magliano, J. P. (2019). The scene perception & event comprehension theory (SPECT) applied to visual narratives. *Topics in cognitive science, 12*(1), 311-351.

Magliano, J. P., Kurby, C. A., Ackerman, T., Garlitch, S. M., & Stewart, J. M. (2020). Lights, camera, action: the role of editing and framing on the processing of filmed events. *Journal of Cognitive Psychology, 32*(5-6), 506-525.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics, 2*(9), 437-442.

Van Dijk, T. A., & Kintsch, W. (1983). Strategies of discourse comprehension.

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

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological science*, *6*(5), 292-297.